CLASS AND CONDITIONAL REASONING
IN CHILDREN AND ADOLESCENTS

by

Elizabeth A. Boyle

A dissertation presented towards the degree of
Doctor of Philosophy in the University of Edinburgh

1984

CONTENTS

ABSTRACT

The development of the ability to comprehend and reason with class and conditional logic statements was examined in the light of Piaget's claim that prior to the age of 11-12 years children are limited to reasoning in terms of classes and relations but from the age of 11-12 years reasoning in terms of propositions becomes possible. Subjects from 5 years to 17.5 years were presented with several different comprehension and inference tasks with class and conditional logic statements. Evidence of differences in the ability of subjects under 12 years to verify class and conditional logic statements was consistent with Piaget's claim that the logical classification operations of the concrete subject enable him to interpret class inclusion statements but that the conditional interpretation of empirical information requires formal operational thinking. No distinction in performance between class and conditional statements was found on tasks which required an understanding of the logical consequences of the inclusion relation with subjects younger than first year secondary performing poorly on both class and conditional versions of an evaluation task and a syllogistic reasoning task. Significant changes in patterns of response at adolescence on the conditional verification task, the evaluation task and the syllogistic reasoning task supported Piaget's contention that there are qualitative changes in reasoning at adolescence although, as in other studies, errors in reasoning by adolesecents indicated that Piaget overestimated the logical abilities of the formal subject and suggested that Piaget's logical model of cognition should be regarded as a model of logical competence.

ACKNOWLEDGEMENTS AND DECLARATION

This research was carried out while the author held a Social Science Research Council Studentship.

My thanks go to members of the Departments of Psychology and Epistemics in Edinburgh University for discussing various problems which occurred during the conceptualization, execution, analysis and interpretation of the work in this thesis and particularly to my supervisors Dr. T. F. Myers and Dr. E. B. Richards.

I am grateful to Lothian Region Education Department for granting permission for access to Blackhall and Davidsons Mains Primary Schools and Broughton High School and The Royal High School in Edinburgh where the research was carried out and also to the teachers and pupils involved for their co-operation.

My thanks are also due to friends and relations, particularly Mrs. Cathie Boyle and Mrs. Caroline Harvey, who have helped to amuse my son Richard while I have been working on the thesis and without whom I could not have completed it. Special thanks are due to my husband Jim for his love, support and encouragement throughout.

I declare that I have composed this thesis myself and that all the work reported in the thesis is my own.

Elizabeth A. Boyle.

GENERAL INTRODUCTION

This thesis is concerned with the acquisition of deductive reasoning ability and the role of logic in explanations of deductive reasoning ability. More specifically it is an investigation of the abilities of children and adolescents from 5 years to 18 years to comprehend and reason with universally quantified statements of the type "All A are B." and the logically equivalent generalized conditional statements of the type: "If x is an A, then x is a B.". The development of the ability to reason with these class and conditional statements is of interest in the light of Piaget's claim that prior to the age of 11-12 years children are limited to reasoning in terms of concrete classes and relations whereas children from 11-12 years onwards can reason in terms of propositions. This claim has been interpreted (Roberge and Paulus, 1971; Ennis, 1975, 1976; Osherson, 1975) as entailing that subjects should be able to reason correctly with arguments stated in the language of class logic before they are able to reason with conditional, or more generally propositional, logic arguments.

Propositionally based theories of reasoning would make no such prediction about the relative difficulty of class and conditional logic problems. On the contrary propositional theories of reasoning propose that the difficulty of a problem is a function of its propositional structure. Since universally quantified inclusion statements like that above are logically equivalent to conditional statements at the level of analysis proposed in propositional theories, these theories would predict that performance on class logic problems should not differ from that on conditional logic problems and the ability to reason with such problems would be predicted to develop synchronously.

Chapter 1, the introduction, is divided into two parts. In part 1 current theories of reasoning are reviewed within the context of recent work of logicians, linguists and psycholinguists and cognitive psychologists who have all been concerned with trying to elucidate different aspects of the complex relationships between logic, language and thinking. The influence of recent research in psycholinguistics and work in formal semantics on the development of theories of reasoning is reviewed. Little consensus of opinion is found concerning the nature of deductive reasoning ability with some theorists advocating a mental logic, the view that reasoning processes accord with logical rules of inference, and others maintaining that psychological aspects of reasoning performance have been neglected by those who concentrate on the logical structure of reasoning problems. Currently the most popular position vis a vis the relationship between logic and thinking is probably the view that logic prescribes a logical competence but that a complete theory of reasoning requires an account of factors which influence performance on reasoning tasks.

Part 2 examines how the ability to make deductive inferences is acquired. Since Piaget's theory has dominated the literature on the development of logical thinking his account of the acquisition of logical thinking and particularly his account of the transition from concrete to formal operational thinking is described in detail. The implications of Piaget's theory with respect to the ability to reason with self-contained linguistic premises is reviewed. Different researchers have interpreted experimental evidence from inference tasks in conflicting ways, as either compatible with or counterevidence to Piaget's theory and some researchers have suggested that Piaget's theory is not relevant to explaining the acquisition of logical reasoning ability with self-contained verbal premises (Ennis, 1975, 1976; Wason, 1977; Braine, 1978; Falmagne, 1975, 1980). It is argued, however, that Piaget did intend his theory to deal with linguistic reasoning although his concentration on the logical structure of problems and failure to consider the plethora of linguistic and task specific factors which influence performance on reasoning tasks make his account of linguistic reasoning incomplete.

The experimental chapters in this thesis examine the evidence for the qualitative changes in reasoning ability at adolescence which Piaget would predict and attempt to identify some of the linguistic and task specific factors which influence response on comprehension and inference tasks involving class and conditional logic statements. The experiments in Chapters 2, 3 and 4 are concerned with the verification of universally quantified and general conditional statements. Chapter 5 discusses the evaluation task, a task requiring the evaluation of exemplars as compatible or incompatible with a general rule while Chapters 6 and 7 are concerned with syllogistic reasoning with class and conditional premises.

CHAPTER 1

PART 1: THEORIES OF DEDUCTIVE REASONING: BACKGROUND AND CONTEXT

The Historical Context of the Relationship between Logic and Thinking

The nature of the relationship between logic and thinking has been a philosophical issue since the days of Plato and Aristotle. Until the late 19th. century most philosophers were of the opinion that the principles of logic were the laws of thought. When Aristotle formulated the principles of traditional formal logic - the rules determining the validity of categorical syllogisms - he believed that the rules he had proposed represented the forms of thought.

Implicit in the title of George Boole's (1854) treatise on the algebra that bears his name (Boolean algebra), "An investigation of the laws of thought." is his view of the nature of the symbolic system he had developed. The laws of logic were, for Boole, the laws of the operation of the human intellect. Had the human intellect operated differently, so too the principles of logic would have been different.

Kant (1885) saw no reason to challenge the traditional view of logic as the science of the necessary laws of thought and Mill (1874) endorsed Archbishop Whateley's definition of logic as the science as well as the art of reasoning. By the science of reasoning he meant "the analysis of mental processes which takes place whenever we reason" (Mill, 1965, p. 2). By the art of reasoning he meant "the rules, grounded on that analysis for conducting the process correctly" (op. cit). In modern terms this definition of reasoning would constitute the proposal that both the competence (the art of reasoning) and performance (the science of reasoning) components of a reasoning model are logically based.

A radical change in the climate of opinion about the relationship between logical principles and thinking was brought about by Frege (Haack, 1978) who claimed that logic and mental processes are unrelated and that logic would exist even if the human mind did not since logical relations are independent of our experience of them. The main objection that Frege had to psychologism - the view that the laws of logic are a function of human cognition - was that mental processes are subjective while logic is objective. For Frege it was not possible that logic was rooted in cognition since no general principles could be derived from something that is subjective. Frege, along with other advocates of antipsychologism including Russell and Wittgenstein, argued that the laws of logic are not the laws of thought but rather that the laws of logic are the laws of necessary relationships between propositions where a proposition is true if it is an accurate representation of external reality.

Another view of the nature of the laws of logic is that the laws of logic are the most general laws of language. One version of this theory is the "linguistic-conventionalist" theory (Mitchell, 1962) such as that of Strawson (1952), which holds that the laws of logic arise out of the general laws governing the conventions of language use. Thus the law of non-contradiction for example would be based on the meaning of the linguistic sign used for negation. The logical positivists also proposed that logic is a language and should be related to a general syntax and semantics; they distinguished between analytic truths, which can be proved only by means of logical laws, and synthetic truths which are established empirically. There are many problems with this view of the nature of logic, particularly in distinguishing analytic from synthetic truths. The logical positivists however were instrumental in the development of formal semantics.

A deeper understanding of the relationships between logic, language and reality has been achieved in recent years due to advances made by logicians and linguists in the study of the semantic analysis of formal systems and the role of formal systems in the analysis of natural language and also because of the revolution in the study of linguistic structure inspired by Chomsky. Proposals like that of the Generative Semanticists that the semantic base of language can be characterised in terms of its logical form obviously imply a closer relationship between logic and cognition than the strong anti-psychologism of Frege although the relevance of much of the logico-linguistic research to psychological questions is still rather unclear.

For psychologists one of the most interesting areas of research in the study of the complex interrelationships between logic, language, thinking and reality has been in explaining how human beings make deductive logical inferences and how they come to acquire this ability. Psychologists have tended to approach the problems surrounding the nature of logical knowledge and the relationship between logic and thinking in a different way from logicians and philosophers. Rather than being concerned with the origins of and nature of formal logic, psychologists have been more concerned with asking whether or not reasoning is logical i.e. whether reasoning is carried out in accordance with the logical principles prescribed by logicians. More recently psychologists have started to appreciate the philosophical problems surrounding the origins and nature of logical systems and consequently have been forced to examine the nature of logical knowledge at a deeper level.

The basic philosophical positions vis à vis the relationship between logic and thinking also have adherents within the literature on the psychology of deductive reasoning. Some psychologists, including advocates of Henle's (1962, 1978) "rationalist" school still adopt the view that the laws of logic describe the way we think. Piaget is also an advocate of the view that formal logic provides a model of logical thinking although Piaget proposes that logical knowledge is constructed by the interiorization of and abstraction from the subject's actions rather than proposing that knowledge of logical principles is innate. Others (Evans, 1982; Pollard, 1982; Johnson-Laird, 1982) have argued that the laws of logic are irrelevant to the way that we think and that concentrating on the logical structure of problems rather than more fundamental aspects of the cognitive processes carried out in solving deductive reasoning problems has obscured rather than clarified the nature of deductive reasoning abilities. However, probably the most acceptable position currently is the view that logic is normative with respect to reasoning i.e. that the validity of an inference can be assessed with respect to formal logic but the mental processes carried out in making the inference do not necessarily correspond to the logical rules of inference (Falmagne, 1980; Cohen, 1981). This could be called a "weak rationalist" position advocated by those who argue that formal logic prescribes a subject's logical competence but that a complete theory of deductive reasoning requires a specification of performance factors affecting the actual conclusions drawn in making deductive inferences.

The following areas of research concerning logic, language and reasoning have been of most obvious relevance in determining the development of theories of reasoning in recent years.

a) Henle's (1962) paper was influential in promoting the rationalist view of the relationship between logic and reasoning. According to this view the reasoning process is logical i.e. the actual mental operations carried out in making an inference can be identified with the rules of logic. This position has inspired much debate in recent years. Critics of the rationalist position have raised several objections to it, notably the problems of errors in reasoning, the effects of content on response and the problem of how a mental logic is acquired.

b) The psycholinguistic research inspired by Chomsky's revolutionary approach to linguistics was a major influence in the development of theories of deductive reasoning with linguistic premises. The aim of the psycholinguistic enterprise was to characterise a) the nature of the cognitive structures and processes underlying the comprehension and production of natural language sentences and b) the relationship between these cognitive structures and operations and grammatical structures and rules. Psychologists interested in reasoning began to draw parallels between the objectives of the psycholinguists and their own objectives in characterising the nature of the representations and structures used in reasoning and the relationship between the representations and operations proposed in psychological theories of reasoning and logical systems.

c) Psychologists (Johnson-Laird, 1983) have recently become increasingly interested in developments in the application of formal systems to the analysis of meaning and inference in natural language - model theoretic semantics. It is important to psychologists to determine whether the formal characterisation of meaning provided by model theoretic semantics is relevant to providing a psychologically realistic theory of meaning and inference. Johnson-Laird (1982, 1983) has recently rejected a simple account of the relationship between formal and psychological semantics but he has obviously been heavily influenced by model theoretic semantics in providing his mental model theory of meaning, representation and inference.

d) The major influence over the past twenty-five years on explanations of the acquisition of logical reasoning ability has been the work of Piaget. In the course of his extensive writings Piaget provided a complex framework describing different stages in the child's acquisiton of knowledge from birth to adolescence. Of particular interest in this thesis is Piaget's account of the concrete and formal operational stages of development and the transition from the former to the latter, particulary in the light of Piaget's claim that for the formal operational thinker "reasoning is nothing more than the propositional calculus itself" (Inhelder and Piaget, 1958, p.305).

In the remainder of Part 1 the research mentioned in (a), (b) and (c) above will be discussed. In Part 2 the acquisition of deductive reasoning ability will be discussed with particular reference to Piaget's account of developmental stages in logical thinking.

Henle and the Rationalist Approach to Reasoning

An article by Henle in 1962 entitled "On the relation between logic and thinking" did much to renew the interest of psychologists in reasoning. Up until then the prevailing climate of opinion amongst psychologists had been that logical principles were not relevant to reasoning. Support for this opinion came from studies such as that of

Bruner, Goodnow and Austin (1956), Lefford (1946) and Morgan and Morton (1944), who found that subjects in reasoning experiments tend to accept conclusions which are empirically reasonable and which accord with their own knowledge, beliefs, experience etc. rather than conclusions which follow validly from the premises. In other words subjects tend to make inferences which are pragmatic or psychologically acceptable rather than logically valid.

The fact that subjects make errors in reasoning was regarded by some, Cohen (1944) and Nagel (1956), as irrefutable evidence that the laws of logic cannot be the laws according to which we think. Henle however pointed out that the existence of errors in reasoning was noted by the traditional philosophers but was not regarded as evidence that reasoning is not logical. Boole (1854), for instance, considered that the existence of fallacies in reasoning was due to interference of other laws with the laws of inference, but did not regard this as evidence that rules of inference did not exist in the subject's mind. Mill (1874) stated "it is scarcely ever possible to affirm that any argument involves a bad syllogism" (Mill, 1874, p. 560). Henle claims that the change from the psychologism of the traditional philosophers to the antipsychologism of the more modern philosophers was due to "an altered intellectual climate rather than any fundamental discoveries about the nature of reasoning" (Henle, 1962, p.94). The change was principally due to the influence of Frege upon ideas about the nature and origins of logic.

Henle reintroduced a rationalist viewpoint - the idea that reasoning is carried out in accordance with the principles of logic - as a respectable position to take vis a vis the relationship between logic and reasoning. Henle claimed that support for this position came from a study of the errors subjects made in evaluating the validity of inferences. Henle identified four different types of error: the first type of error was an error of non-acceptance of the logical task, similar to the type of responses reported by Morgan and Morton, (1944), Bruner et al (1956): responses of this type were based upon personal beliefs, opinions etc. rather than upon purely logical considerations. Where the subject did accept the logical task Henle identified three different types of error based on the subjects' misinterpretation of the premises: subjects added in extra premises, they omitted to consider premises and they interpreted the premises in idiosyncratic ways thereby changing the intended meaning of the premises. Henle concluded that her subjects' inferences were always logically valid if the validity of an inference was judged with respect to the subjects' interpretation of the premises within the context of the problem as a whole, rather than being judged with respect to the interpretation intended by the experimenter. This rationalist hypothesis, that reasoning is carried out by means of logical inference on misinterpreted premises, has inspired much research on reasoning in the past twenty years. As recently as 1978, in the introduction to Revlin and Mayer's "Human Reasoning" Henle reaffirmed her commitment to the rationalist position: "I have never found errors which could unambiguously be attributed to faulty reasoning".

One reason for the popularity of Henle's rationalist hypothesis is that it gets out of the problem which faces any empirical investigation of reasoning of the circular relationship between logicality and understanding. The problem, which was described by Smedslund (1970), is that in order to say whether a subject is responding logically in a reasoning task one must assume that he has understood the task instructions and premises. On the other hand if one wants to assess a subject's understanding of a statement one must assume that the subject is responding logically. Smedslund originally described the problem as a problem for Piagetian psychology in which assessment of the child's logical operational level is based upon the assumption that the child has understood terms and instructions in the task; however the problem is a general one and exists for any psychological explanation of responses on reasoning tasks.

With the growth of interest in psycholinguistics in the late 1960s and early 1970s new paradigms were being sought for investigating how children and adults understand sentences. It was realized that the rationalist hypothesis offered a way of studying sentence comprehension. Since errors in reasoning were attributable to misinterpretation of the premises, subjects' responses on reasoning tasks would reflect their understanding of the premises. Although this solution seems attractive to the psycholinguist interested in sentence comprehension, it would leave the researcher in reasoning in the paradoxical position of investigating a subject's logical ability but presupposing that he behaves logically!

It seems that much of the confusion surrounding this debate about logicality and understanding occurs due to the misuse of the term 'logical'. Smedslund (1977) describes how, following his recognition of the circular relation between logicality and understanding, he realised that the subject must always be regarded as rational given his interpretation of the situation. Smedslund claims that "it is possible and intuitively obvious to presuppose logicality whereas it is clearly absurd to presuppose understanding". It is not clear what Smedslund means by 'logicality' but if he means that we should regard reasoning processes as following valid logical principles it is by no means obvious that we should assume logicality. It is possible that he means something weaker than this: that the subject should be regarded as generating responses which can be seen to accord with the responses that logic would prescribe given his interpretation of the premises. It seems that Smedslund is in fact saying something even weaker than this - that the subject should be presumed to behave logically given his understanding of the problem and given his level of logical development, where 'logically' seems to mean 'sensibly' and in accordance with a limited system of rule governed behaviour.

One of the major problems with the rationalist theory is its untestability. If the subject is presumed to make logical inferences based upon his own personal understanding of the premises we would have to know how the subject understood the premises in order to affirm that his reasoning was logical. The subject's comprehension of the premises would have to be determined independently of the inferences he makes; however the subject's comprehension of the premises is inferred from the inferences he makes, and the logical nature of the inferences is presupposed. Looking at this in another way it is always possible to infer some interpretation of the premises which would logically imply any given conclusion. There would be no point in simply asking the subject how he interpreted the premises because introspections of this type are notoriously unreliable (Evans, 1980, 1982).

Criticisms of Rationalism

One of the most severe critics of the rationalist position is Evans (1978, 1980, 1982). Evans (1980) argues that if subjects do use logical rules of inference in reasoning, they should show a consistent pattern of reasoning both within and between tasks. A subject who interprets a conditional statement as a biconditional in one task should also interpret it in this way on other tasks. Evans (1978) found however that subjects who interpret the conditional as an implication (conditional) on an evaluation task, such as that described by Johnson-Laird and Tagart (1969), apparently interpret it as an equivalence (biconditional) on a conditional syllogism task. In addition to this, responses on Wason's selection task reflect neither a conditional nor a biconditional interpretation. The rationalists would not accept this difference in interpretation between tasks as evidence that reasoning is not logical; they would argue that reasoning is logical but task specific factors influence the interpretation of the conditional. Between task differences in response can be explained if it is assumed that reasoning is logical given the different interpretations of the premises in the different tasks. The non-rationalist would object that it is incumbent upon the rationalist to specify how the task specific factors determine the interpretation of the conditional.

A further problem which Evans poses for the rationalist hypothesis is the existence of non-logical response biases in reasoning which in many cases provide better explanations of errors than the Henle hypothesis. One of the earliest examples of a non-logical theory was the atmosphere hypothesis of Woodworth and Sells (1935). Evans calls the theory non-logical because response was explained not in terms of the logical form of the premises but in terms of a global impression of the "atmosphere" of the premises. The atmosphere hypothesis predicts that whenever premises contain negative or existential premises so too the most frequently drawn conclusions will contain negative or existential premises.

Further examples of non-logical response biases which Evans investigated include 1) negative conclusion bias and 2) matching response bias. Evans reported these biases on a variety of reasoning tasks including categorical and conditional syllogisms (Evans, 1977), Wason's selection task (Evans and Lynch, 1973) and a task in which subjects were required to provide verifying and falsifying instances of a conditional rule (Evans, 1972).

The rationalist model of reasoning holds that inferences are carried out in terms of logical rules of inference and that these rules of inference operate on problems with the same logical stucture regardless of the content of the problem. There are however many studies in the literature which demonstrate that the content of the problem has a significant effect on the conclusions drawn. One of the most impressive and well-researched examples of this occurs in Wason's selection task.

In the selection task the subject is presented with four cards showing 'A', 'D', '4' and '7' and he is told that each card has a letter on one side a number on the other. He is then given a rule about the four cards in front of him e.g.

"If a card has a vowel on one side, it has an even number on the other side."

His task is to say which of the cards he would need to turn over in order to find out whether the rule is true or false. Wason and Johnson-Laird regarded the task as very similar to the formal operational hypothesis testing tasks described by Inhelder and Piaget (1958) and consequently they expected that adults should be able to solve it. However only 5 out of 128 adult subjects in four studies reported by Wason and Johnson-Laird gave the correct response on the task, which in the example above was 'A' and '7'. Most subjects chose 'A' alone or 'A' and '4'. In other words most subjects fail to appreciate the importance of choosing '7' as a possible falsifying instance of the general rule. Wason and Johnson-Laird argued that this result raised considerable problems for Piaget's account of formal operational thinking since, according to Piaget, the formal operational subject tests different hypotheses by actively looking for falsifying exemplars to particular hypotheses.

Even more difficult to explain was the dramatic improvement in performance on the selection task when the problem was presented with realistic content. Wason and Shapiro (1971) presented their subjects with the rule:

"Every time I go to Manchester I travel by train."

and cards showing 'car' 'train' 'Manchester' 'Sheffield'. 10 out of 16 subjects identified the cards 'Manchester' and 'car' as the correct cards to turn over in this study compared with only 2 out of 16 in the abstract version of the problem. In a further study by Johnson-Laird, Legrenzi and Legrenzi (1972) with realistic content, 20/24 subjects correctly identified sealed letters and letters with 40 lire stamps (not 50 lire stamps) as the letters to turn over in order to say whether or not it was true that:

"If a letter is sealed it has a 50 lire stamp on it."

However correct response on this realistic problem did not transfer to the abstract problem. Much research has been devoted to trying to establish the origin of this so-called 'thematic materials' effect with the selection task, with Gilhooly and Falconer (1974) attributing the effect to the concrete/abstract nature of the premises and Bracewell and Hidi (1974) arguing that the relation between the terms was a more important determiner of performance. Van Duyne (1974, 1976) found that both factors contributed to the thematic materials effect.

The superior performance found on this and other reasoning problems with realistic content and the failure to transfer to problems with abstract content is difficult to reconcile with a rationalist model although again the rationalist would argue that different types of content influence the interpretation of the premises (Staudenmayer, 1975). Cohen (1981) argued that the selection task is a cognitive illusion evoked by the strange task requirements. He regarded the results of the various manipulations of the selection task as providing important data concerning the relationship between the underlying logical competence and performance on the task.

Finocchiaro (1980) has recently argued that the facilitative effect of concrete content in the selection task can be given a formal explanation. He argues that in treating the rules in their selection task as propositional conditionals Wason and Johnson-Laird provide an oversimplistic analysis since the rules are in fact universal generalizations of conditionals involving quantifiers. Furthermore, under Finocchiaro's analysis, the abstract rules are much more logically complex than the concrete rules and, since logical complexity presumably correlates with problem difficulty, problems involving abstract rules would be predicted to be more difficult than problems involving concrete rules. Finocchiaro argues that the predicate logic representation of the concrete rule: "Every time I travel to Manchester I go by train." is:

$\forall(x)(Mx \rightarrow Tx)$ where Mx: x is a journey to Manchester.

Tx: x is a journey by train.

whereas the formalization of the abstract rule: "If a card has a vowel on one side it has an even number on the other side." is:

$(x)(y)(z)(Cx \,\&\, Pxy \,\&\, Syx \,\&\, y \neq z \,\&\, Szx \rightarrow Qxz)$

where:

Cx: x is a card

Pxy: x has a vowel on side y

Syx: y is a side of card x

Qxz: x as an even number on side z

Although Finocchiaro indicates that it is possible to give a formal explanation of the facilitative effects of concrete content he argues that the reasoning processes which would be required to solve the task formally would be so complex that the formal explanation is implausible.

A recent attempt has been made by Pollard (1982) to provide a more coherent, integrated explanation of non-logical response biases and the facilitative effect of concrete content in terms of Tversky and Kahneman's availability heuristic. The marked facilitative effect that concrete content has on a variety of responses in reasoning problems and the apparent inability of subjects to ignore the truth value and content of a conclusion, even when explicitly asked to do so, illustrates, according to Pollard, that subjects are more concerned in real life with making correct decisions than they are in making inferences that are logically valid. The knowledge that a conclusion is a true statement is generally more useful to a subject than the knowledge that it is a logically valid conclusion. Subjects base their responses to reasoning problems on "available cues" i.e. the information relevant to the content of a problem that can be retrieved from memory.

Non-logical response biases and content effects do not pose such a problem for the more recent versions of the rationalist hypothesis as they did for the older, more extreme versions of the hypothesis. According to the more recent versions of the rationalist hypothesis the subject may have an underlying logical competence which constitutes his knowledge of logical principles but in using and understanding logical arguments "performance factors" would have to be considered. Non-logical response biases would be regarded as a class of performance factor. Although a subject may possess a logical competence he may fail to apply logical rules in reasoning for a variety of reasons e.g. the content of the problem may influence his response; he may decide that an argument is acceptable even although it is not deductively valid; he may realize that he should apply certain inference rules but lack the motivation to apply them. Errors in performance on a logical task would not necessarily indicate that the subject could not, in other circumstances, understand the logical principle involved but would indicate instead "some malfunction of an information processing mechanism" as Cohen (Cohen, 1981 p.317) puts it. Instead of regarding content effects as evidence that inference is not logical the competence/performance models allow that the individual has an underlying competence and that the psychologist's task is to identify the factors which determine how the subject interprets the task.

The distinction between competence and performance will be recognised as similar to that which Chomsky (1965) introduced into linguistics. Just as the speaker's knowledge of language constitutes his linguistic competence so the reasoner's knowledge of logical principles constitutes his logical competence, and just as linguistic performance is concerned with the principles governing language understanding and production so logical performance is concerned with the principles which govern the production and comprehension of logical arguments. The older rationalist hypothesis can be compared with an integral model of grammar, where the logical (grammatical) rules form an integral part of a theory of performance (Steinberg, 1982). The more modern versions of the rationalist theory propose that competence and performance are two separate components of a theory of reasoning and are more like componential models of grammar: logical rules constitute a component of the theory of reasoning which interact with other components to determine performance on reasoning tasks.

One of the major problems with competence theories, as Evans (1980) points out, lies in their untestability. The competence component is frequently inferred from inferences which subjects have made or which they find acceptable. The competence component is, in other words, inferred from subjects' performance on reasoning problems. However if both competence and performance components are to be specified in a model of reasoning they should be characterised independently. The competence component is frequently given an independent characterisation as some formal logical system. This approach is also problematic however in that it is not clear which logical system should be used as a competence model. The unsuitability of standard propositional logic as a model of competence in reasoning with natural language premises has been stressed by Wason and Johnson-Laird (1972), Johnson-Laird (1975), Osherson (1975), Braine (1978). Formal logic incorporates paradoxes which reasoners would find unacceptable, including the paradoxes of material implication and disjunction. Different inferences will be valid in different logical systems and so whether a particular inference is judged as valid or as invalid depends upon the logical system adopted as a competence model.

Another problem with formal logic as a model of reasoning is that logical systems are usually formalized axiomatically i.e. as a set of axioms with a small number of rules of inference. This type of formalization seems inappropriate as a model of human inference since human beings are not normally concerned with proving logical truths but rather they are concerned with establishing meaningful/pragmatic connections between statements. Gentzen proposed a system of natural deduction, the explicit aim of which was to "set up a formalism that reflects as accurately as possible the actual reasoning involved in mathematical proofs" (Gentzen, 1964, p. 291). The system was a formalization of standard logic which dispensed with axioms and consisted only of inference rule schemata. An inference rule schema defines rules of inference by specifying their form and a rule of inference consists of an instruction or warranty for concluding y given x, e.g. an inference rule schema for conjunction would be:

$$\frac{p \quad q}{p.q}$$

The relevance of the natural deduction system as a possible model of human inference was discussed by Johnson-Laird (1975), Osherson (1975), Braine (1978). Inference rule schemata have the advantage as psychological models that they do not incorporate the paradoxes of standard propositional logic since they have an inherent directionality.

Osherson (1974, 1975) attempted to correlate the difficulty experienced by both adults and children in attempting to solve certain logical problems, assessed in terms of both latency of solution and error rates, with the difficulty of the problems in terms of the number of steps involved in different algorithmic proof procedures in propositional logic: these proof procedures included the truth table method, reduction to disjunctive or conjunctive normal form, and methods for establishing falsifying truth value assignments. Osherson found no correlation between formula difficulty and any of these proof procedures. Osherson then looked at the correlation between the experienced difficulty and the number of inference rules in a natural deduction procedure. Osherson found a correlation of 0.6 between experienced difficulty of a problem and the number of inference rules proposed for a successful solution to the problem. This rose to 0.8 when appropriate differential weightings were given to operations involving disjunctive and conjunctive operations in order to account for the difficulty of disjunction compared with conjunction (Bourne and Banion, 1971; Neimark and Slotnick, 1970). Despite this success Osherson acknowledges the limitations of his natural deduction system in that it only deals with a subset of propositional logical truths and not at all with predicate logic.

Johnson-Laird (1975) also proposed a natural deduction system for propositional inference as a model of competence in propositional reasoning. He argued that it was an empirical problem to determine which inference rule schemata are psychologically basic and the matter can only be resolved by determining which inferences mature but logically naive people find acceptable.

The Influence of Psycholinguistic Theories on Theories of Reasoning

The adoption of the competence/performance distinction by psychologists interested in reasoning, or psychologicians as Osherson calls them, was not the only influence of psycholinguistic theories on theories of reasoning in the 1960s and 1970s. Many of the issues which concerned the linguists and psycholinguists who were studying the mental representations and processes underlying the comprehension and production of natural language sentences were also of interest to psychologists attempting to characterise the mental structures and operations used in making inferences.

One of the major arguments in linguistic theory following the publication of Chomsky's "Aspects of a Theory of Syntax" in 1965 concerned the existence of the level of deep structure as defined by Chomsky. The standard generative transformational grammar proposed in the Aspects model was a sytactically based grammar. The language was formally characterized in terms of two levels of syntactic structure, the deep structure and the surface structure, and a sequence of operations called transformations relating these levels. The deep structure level was the input to the semantic component and all semantic information relevant to the interpretation of a sentence was held to be present in the deep structure (See Figure 1.1).

Syntax and semantics in the Aspects model were separate both in the sense that syntactic categories were not semantically based and also in the sense that the transformational sequence from deep structure to surface structure was syntactically based. In a similar way rationalist models of reasoning proposed that the semantic

FIGURE 1.1

CHOMSKY'S STANDARD THEORY OF GRAMMAR

$$
\begin{tikzcd}[column sep=small]
 & \text{BASE RULES/LEXICON} \arrow[d] & \\
 & \text{DEEP STRUCTURE} \arrow[dl] \arrow[dr] & \\
\text{TRANSFORMATIONAL RULES} \arrow[d] & & \text{SEMANTIC RULES} \arrow[d] \\
\text{SURFACE STRUCTURE} \arrow[d] & & \text{SEMANTIC REPRESENTATION} \\
\text{PHONOLOGICAL REPRESENTATION} & &
\end{tikzcd}
$$

FIGURE 1.2

GENERATIVE SEMANTICS MODEL OF GRAMMER

$$
\begin{tikzcd}
\text{SEMANTIC REPRESENTATION} \arrow[d] \\
\text{TRANSFORMATION RULES AND LEXICON} \arrow[d] \\
\text{SURFACE STRUCTURE} \arrow[d] \\
\text{PHONOLOGICAL RULES} \arrow[d] \\
\text{PHONETIC REPRESENTATION}
\end{tikzcd}
$$

interpretation and logical inference components were separate with semantic interpretation preceding logical inference and inference rules operating in a strictly formal way upon the well formed formulae specified in the semantic encoding.

Chomsky's description of a grammar in terms of linguistic structures and rules and his conviction that conclusions about the nature of the human mind could be drawn from generalisations about the nature of linguistic structure led to an intensive effort by psychologists in the 1960s and 1970s to establish whether the structures and operations he proposed did correspond to identifiable cognitive structures and operations used in language comprehension and production. In their review of the literature in 1974, Fodor, Bever and Garrett concluded that while the evidence supported the view that deep and surface syntactic levels corresponded to pschologically real levels of encoding it did not provide support for transformations as cognitive operations carried out during sentence comprehension.

Clark (1969) was one of the first to propose that the kinds of mental strategies and processes proposed in theories of language comprehension and production can also be applied to theories of deductive reasoning. Clark's linguistic model of reasoning proposed that the premises of the reasoning problem are encoded using the same processes used in sentence comprehension. Clark proposed that the deep structure level of a sentence, the abstract structure accessed in interpreting the sentence, contains the information relevant to making deductive inferences with the sentence. He showed that three principles which were related to the deep structure rather than the surface structure of the sentence - the principle of the primacy of functional relations, the principle of lexical marking and the principle of congruence - were good predictors of the mean solution times for two and three term series problems. This conception of deductive reasoning provided the basis for a more general theory of deductive reasoning as closely integrated with language comprehension in contrast with previous accounts of deductive inference which tended to be specific to particular reasoning tasks, e.g three term series problems (Hunter, 1957; Donaldson, 1963; Handel, De Soto and London, 1968; Huttenlocher, 1968) and syllogistic inference (Woodworth and Sells, 1936; Chapman and Chapman, 1959).

Subsequent research, linguistic and psycholinguistic, questioned the status of the level of deep structure defined in Aspects as the level from which the semantic representation of a sentence is specified. It was challenged by the interpretive semanticists, including Jackendoff (1971) and Chomsky himself (1971), as too deep. They claimed that certain aspects of sentence meaning including stress placement and scope and quantification properties should be defined at the level of surface structure.

The Generative Semanticists including McCawley, Lakoff, Ross, Postal (Lakoff, 1971) on the other hand, claimed that the level of deep structure specified in Aspects was too shallow and was, in fact, identical to the level of semantic representation. They challenged both the primacy and autonomy of syntax, arguing that syntactic categories are semantically motivated and that semantics should play a primary or generative role in grammar. In addition they argued that there is no exclusively syntactic derivational sequence mapping the ultimate underlying structures on to the surface structures. Rather the grammar would consist of a interacting sequence of syntactic (transformational) and semantic (lexical insertion) rules which would relate the ultimate underlying structures, the semantic representations, to the surface structures (see Figure 1.2).

Some Generative Semanticists (Lakoff, 1970, 1971) identified the level of semantic representation with the logical form of the sentence although others (Seuren, 1972) were more cautious, arguing that it was premature to specify the nature of semantic representations.

The Generative Semanticists' identification of deep grammatical structure with logical form was of interest to logicians interested in characterizing the logical structure of natural language (Harman, 1970) since it suggested that the concerns of linguists and logicians in analysing natural language were more similar than had sometimes been supposed. Cresswell (1973) expressed surprise that the Generative Semanticists chose first order logic rather than a more complex formal language as a base language for analysing linguistic structure and described much richer formal languages which he felt were more appropriate to the logical analysis of natural language.

The grammar advocated by the Generative Semanticists was also of interest to psychologicians since it could be regarded as an integral model of language comprehension or, more accurately, language production. Many of the Generative Semanticists regarded the grammar (See Figure 1.2) as a model of psychologically real structures and processes carried out in actually understanding and producing sentences. The semantic representations which the Generative Semanticists proposed as psychologically real levels of semantic representation produced or accessed in producing or comprehending a sentence would be identical to the logical form of the sentence, the level from which valid inferences could be specified. Thus the semantic representations proposed by the Generative Semanticists were just the representations which psychologicians wanted to characterise. The arguments underlying the Generative Semanticists' challenge to the primacy and autonomy of syntax in Chomsky's standard theory have parallels in the arguments put forward to challenge the independence of the semantic interpretation and logical inference components proposed in rationalist models of reasoning. Just as the Aspects' model comprising a purely syntactic sequence of rules joining deep and surface structures was rejected by the Generative Semanticists, so too psychologicians found evidence that the semantic interpretation of premises in a reasoning problem cannot be separated from the inferences that the subject makes. The results of many studies show that the inferences subjects make in reasoning tasks cannot be explained in terms of the application of standard logical rules of inference upon a formal representation of the premises. Rather response depends upon the content and context of the problem, presuppositions connected with the premises etc. (Staudenmayer, 1975; Revlis, 1975). Falmagne (1975) argues that these results show that the subject is aware that different situations have different logical properties and that the encoding of the premises incorporates implicit logical knowledge. Staudenmayer argues from a rationalist position that the context, presuppositions etc. influence the subject's interpretation of the premises but that inference is logical, carried out on the individual subject's complete interpretation of the premises.

Propositional Representation

Following the recognition of the problems associated with deep structure, much research in linguistics and psycholinguistics was directed towards characterising psychologically realistic grammars and towards more general problems in determining how sentences and texts are understood and how linguistic information is represented and stored in memory (Kintsch, 1974). It appeared that while the actual words used in the linguistic expression of the input sentences were poorly remembered in recall and recognition tasks the overall meaning, or gist, of the sentences or text was remembered well (Sachs, 1967, 1974; Johnson-Laird et al, 1973; Ratcliff and McKoon, 1978; Brewer and Lichtenstein, 1975; Monaco, 1975). Information from text is apparently represented in an abstract form corresponding to the "meaning" of the text which psychologists have called a propositional representation. The term 'propositional representation' as used by psychologists refers to an abstract mental representation of linguistic or non-linguistic information. Propositional representation of non-linguistic information requires that the information be interpreted in terms of a verbal description. As Johnson-Laird (1983, p.155) puts it: "A propositional representation is a mental representation of a verbally expressible proposition". Propositional representations use propositions as units of meaning which can nevertheless be decomposed into component elements which reflect intrapropositional structure. A proposition is essentially an abstract statement about properties or relations between entities, for example "Jim snores" expresses the proposition that the entity named "Jim" has the property of "snoring". Similarly "Jim rang the bell" expresses the proposition that the relationship between "Jim" and "the bell" is one of "ringing". Propositional representations as representations of the "meaning" of sentences are similar to, but not the same as, the deep structure level in the Aspects Model, since the deep structure level was the level from which the semantic interpretation of a sentence was defined. Since the level of semantic representation defined in Generative Semantics is identical to the level of logical form and since this is usually a quantificational logical form, the semantic representations in Generative Semantics are what psychologists would call propositional representations.

Different theorists have used different notations to express propositional representations but in general propositional representations are equivalent to predicate calculus formulae. Kintsch calls the component elements of the proposition 'word concepts' a term which covers both predicates (properties or relations) and arguments. Clark and Chase (1971) also used a linear notation for propositional representations while others use semantic network notations with nodes and relations (Anderson, 1976, 1978; Norman and Rumelhart, 1975). Although the different notations contain the same information, they have different psychological commitments: relationships in the network notation, for instance, are more explicit. Another difference between the notations is in their commitment to semantic decomposition. Norman and Rumelhart (1975) decomposed words into semantic or conceptual primitives while Anderson and Bower (1973) and Kintsch (1974) did not break down the representation into more primitive meaning components.

Evidence that subjects use propositional representations in comprehending sentences comes not just from studies of memory of information from the text but also from chronometric studies. In sentence verification studies in which subjects had to decide whether or not a simple sentence was an accurate description of a picture (Clark and Chase, 1972; Carpenter and Just, 1975) the sentence verification time was found to depend upon the extent of propositional embedding in the sentence i.e. the number of constituent operations involved in verifying the sentence, including negation and falsity operations. The sentence verification task also correlates reasonably well with measures of general verbal comprehension. These results were understood to indicate either that the information from sentences and pictures is encoded into a common representational format - a linguistic/propositional format-or that translation from propositional to other kinds of mental representation is relatively simple.

Propositional models of representation emphasise similarities between models of comprehension and reasoning since the semantic representations which subjects access in comprehending sentences - the propositional representations - are identical to the logical forms of sentences and the logical form of the sentence is the level from which logical inferences can be specified.

Martin (1981) has recently questioned the adequacy and generality of semantic representations of the kind proposed by proponents of propositional representational systems which advocate an underlying predicate calculus format. Two of the main hypotheses about semantic representations, the complexity hypothesis and the distance hypothesis, make predictions about the relative difficulty involved in comprehending sentences which are not substantiated if semantic representations are quantificational structures. The complexity hypothesis predicts that the more complex the semantic representation of a sentence the greater the amount of processing and consequently the longer the time required to comprehend a sentence. The distance hypothesis predicts that the greater the disparity between the surface form and the semantic representation of a sentence the greater the difficulty of the sentence as measured by the comprehension time. Although predicate calculus representations may be adequate for representing the very simple sentences (or more accurately non-sentences) used by Clark and Chase e.g."Star is above plus.", more complex sentences involving quantifiers and even determiners etc. prove to be problematic. For example the predicate calculus notation for sentences involving the definite article e.g. "The man is fat" is more complex than that for sentences involving the indefinite article e.g. "A man is fat.", the former representation being $x(Mx \mathbin{\&} Fx \mathbin{\&} \forall y \, (My \rightarrow x=y))$ and the latter $x(Mx \mathbin{\&} Fx)$. Both the complexity and distance hypotheses would predict that sentences with the definite article should be more difficult to comprehend than sentences with the indefinite article, a prediction which is not substantiated by the data. In attempting to give a rough quantitative estimate of the degree of difficulty predicted by quantificational models of comprehension Martin proposed that the difficulty of a sentence is a function of its complexity measured in terms of the number of predicates, connectives and quantifiers in the semantic representation of the sentence: this would lead to a prediction that, for example, sentences like "Exactly ten glasses are broken." should be around ten times as difficult to understand as "Exactly two glasses are broken." which is clearly an implausible prediction of any realistic theory of sentence comprehension. Martin claims that this problem exists for any psychological theory of sentence comprehension which proposes that the semantic representation of a sentence is like a predicate calculus notation. Kintsch does not actually treat quantifiers in the standard way but rather regards them as predicates but this causes problems in interpretation of scope and other properties of quantification.

Comprehension and Inference

Despite certain limitations, recall and recognition studies in the

1970s provided an increasingly detailed picture of the characteristics of the semantic representations constructed in the comprehension of sentences and texts. It was becoming increasingly obvious that subjects did not simply construct representations of the information explicitly expressed in the text. Evidence accumulated from both memory studies (Johnson, Bransford and Solomon, 1973) and on-line reading studies (Haviland and Clark, 1974; Clark, 1977) that in constructing a coherent integrated representation of the text, subjects are continually making simple inferences. Johnson et al for example showed in a sentence recognition study that subjects frequently mistakenly identify as sentences that they have heard before, sentences which are merely inferrable from the original sentences. Similarly Keenan and Kintsch (Kintsch, 1974) showed that subjects could not distinguish after fifteen minutes whether a sentence (a) had been explicitly expressed (b) in a sentence or merely implied (c):

a) The cigarette started a fire.

b) A carelessly discarded cigarette started a fire. The fire destroyed many acres of virgin forest.

c) A burning cigarette was carelessly discarded. The fire destroyed many acres of virgin forest.

On-line reading studies were important in demonstrating that inferences are made at the stage of comprehension of the text. Haviland and Clark (1974) and Clark (1977) for instance showed that, in comprehending text, subjects are continually trying to establish identity of reference between current entities and antecedent anaphors e.g. in attempting to understand: "Arthur sold Jim a brand new hatchback. The car broke down on its first trip." the subject will try to establish an anaphor for "car". Since there is no prior reference to "car" the subject will make what Haviland and Clark called a 'bridging inference' to infer an anaphor, which in this case is obviously "hatchback". This inferential process was shown by Haviland and Clark to take a measurable amount of time.

Johnson-Laird and Wason (1977) and Johnson-Laird (1982) distinguish between implicit and explicit inferences in reasoning. The former are smooth, unconscious and automatic while the latter require conscious, deliberate effort. Implicit inferences are inferences made in comprehending text and are pragmatic or plausible rather than deductive. The bridging inferences described above are just one example of these implicit inferences. Johnson-Laird proposes that implicit inferences are necessary for comprehension of a text. According to the extreme rationalist position, deductive inferences are made following comprehension of a text (premises). It is tempting to suppose that the distinction between implicit and explicit inference can be drawn in terms of whether the inference is necessary for comprehension (Johnson-Laird's implicit inferences) or dependent upon comprehension (Johnson-Laird's explicit inferences). Harris and Monaco (1978) however made very different claims regarding logical and pragmatic inference where their logical and pragmatic inferences seem to be roughly comparable to Johnson-Laird's explicit and implicit inferences respectively. Logical implication is that which is logically and necessarily implied by an utterance and pragmatic implication is that which is strongly suggested by an utterance but not directly asserted or logically implied. Harris and Monaco claim that "Unlike logical implications however pragmatic implications do not have to be understood to meaningfully comprehend a sentence.".

Harris and Monaco apparently suggest that the logical implications of a sentence (as well as the directly asserted meaning of a sentence) have to be understood for the sentence to be adequately understood whereas pragmatic implications do not have to be understood for meaningful comprehension but will modify the perceived meaning of the sentence if they are understood.

Why do Harris and Monaco claim that logical implications have to be understood to "meaningfully" comprehend a sentence, while rationalist models of reasoning argue that the logical inference and semantic interpretation stages are separate with logical inference dependent upon interpretation? It is relevant at this point to clarify the difference between logical inference and logical implication, terms which even logicians sometimes use carelessly. Logical implication is a relation which holds between two statements (sometimes defined over propositions) A and B, such that if A logically implies B, B cannot be false when A is true. Logical inference is an operation carried out by a person in deriving new information (a conclusion) from information already known (the premises). Bradley and Swartz (1979, p. 194) define inference as follows:

"an act or series of acts of reasoning which persons perform when, from the truth of a propos ition or set of propositions, P, they conclude the truth of a proposition or set of propositions, Q."

Implication then is a relation between statements whereas inference corresponds to the mental processes carried out in drawing a conclusion from premises. The same sort of distinction can be made between pragmatic implication as a relation between sentences and pragmatic inference as mental processes carried out in drawing a pragmatic conclusion.

Keenan (1978) argues that Harris and Monaco make the typical mistake of psychologists who adopt formal systems as psychological models by assuming that the formal operations entail identical cognitive operations, in this case assuming that if a sentence X implies a sentence Y then the subject will infer Y from X. Keenan rejects Harris and Monaco's claim that encoding a sentence must involve encoding its logical implications. The implausibilty of this claim is evident when it is considered that there are an infinite number of valid implications which follow from any given proposition, for instance p implies p or q, p or q or r, $-(-p)$, $p(q \rightarrow p)$ etc. Arguing that in order to understand a sentence p one has to process all possible logical implications is clearly psychologically unrealistic. Harris and Monaco (1978) subsequently modify this claim to arguing that logical implications are necessarily comprehended "insofar as they are necessary for an adequate comprehension".

Harris and Monaco distinguish between logical and pragmatic implication in terms of sentence structure and seem to want to argue that there are also differences in how the two different types of inference are comprehended in terms of the necessity of the inference for adequate comprehension. The results of their experiments however do not support this view and they conclude that "that which is pragmatically implied is at times functionally equivalent to that which is directly asserted.". Keenan has argued that all meaning whether asserted or implied is derived and that Harris and Monaco's distinction between logical and pragmatic implication and even between asserted and implied information is not a psychologically useful distinction since subjects will differ in the inferences they make according to their background knowledge.

In their discussion paper Harris and Monaco offer a more flexible approach to comprehension and inference in terms of levels of processing. They argue that comprehension of a statement may vary from shallow, in which case only directly asserted information would be comprehended, to increasingly deep levels of comprehension for which increasingly large numbers of logical and pragmatic inferences would be made. The depth of processing of a sentence would depend very largely on the relevant knowledge which a subject brings to bear on the sentence. As an example Keenan's sentence: "Energy is equal to mass times the speed of light squared." will be understood at a high level by a physicist who will understand many logical and pragmatic implications of the sentence but probably only the directly asserted information will be understood by the layman.

The point of the preceding discussion has been to bring out the uncertainties surrounding the semantic representations proposed in accounts of sentence comprehension. Should logical inferences be incorporated into the semantic representation of a sentence? Is the distinction between inferences which are logically valid and those which are merely probable a psychologically useful distinction to make? Although psychologists acknowledge the need to account for a plethora of inferences in natural language which are not strictly logical but which human beings find acceptable, most accounts of inferential ability to date have concentrated on the inferences prescribed by standard logical systems. Most psychologists would maintain a distinction between logical and pragmatic inference but acknowledge the need for further study of similarities and differences between logical and pragmatic inferences.

Alternatives to Propositional Representation

Although many cognitive psychologists argue that all information is represented in the mind in a propositional format (Pylyshyn, 1981) this is not the view of all those interested in representation. There are many psychological experiments which purport to demonstrate that there are two different kinds of functional representation of information - propositional and analogical or imaginal (Brooks, 1968; Pavio, 1975; Shephard and Metzler, 1971; Cooper,1975; Kosslyn, 1975; McLeod et al). Whereas information is represented propositionally in an abstract interpreted form as relationships between entities, an analogical/imaginal representation is an accurate image of the original stimulus and mirrors properties of objects and relationships in the real world.

Although there is no doubt that images actually exist there is controversy over the status of images. Some researchers (Anderson and Bower, 1975; Pylyshyn, 1973; Kieras, 1978) have claimed that images are not functional in representation but are merely a by-product of representation. They argue that information which can be represented imaginally can equally well be represented propositionally. Those who favour images in addition to propositional representations as a form of mental representation usually argue that analogical representations are more suitable than propositional representations for certain kinds of mental operation. These "dual code" theorists have produced experimental evidence of selective interference of type of material - verbal or visual - (and hence preferred representation) with the response required (Brooks, 1968). Cooper (1975), Shephard (1975) and Kossylyn (1975) have also argued that explanations of the abilities of subjects to mentally rotate and transform shapes in terms of analogical/imaginal representations are more plausible and succint than explanations in terms of propositional representations.

Further support for the dual code hypothesis also comes from the data of McLeod et al (1978) and Matthews et al (1979). They found in a verification task that the propositional verification model which proposes that verification time depends upon the extent of propositional embedding provides a good fit when data are collapsed across subjects. When an attempt was made to fit the data of individual subjects to the verification model the fit for some subjects was found to be very good but for others it was very poor. McLeod et al explained the difference between the two groups in terms of different verification strategies - linguistic and spatial - used by the subjects in the two different groups.

McLeod et al devised a task which gave a measure of both sentence comprehension time and sentence verification time. The group of subjects who were inferred to use a propositional/linguistic strategy spent a comparatively short time in the comprehension stage (1652 ms.) compared with those inferred to use a spatial strategy (2579 ms.) and a comparatively long time in the verification stage (1210 ms.) compared with the spatial group (651 ms.). The different comprehension times were understood as showing that the linguistic group encoded the sentence propositionally while the spatial group encoded the sentence imaginally. Thus the verification times for the spatial group on presentation of the picture were considerably shorter than the verification times for the linguistic group. Only for subjects inferred to use a linguistic strategy was the mean sentence verification time a function of the number of propositional embeddings in the sentence while the verification times of those inferred to use a spatial/imaginal strategy did not change as a function of propositional embedding.

Mani and Johnson-Laird (1982) present evidence for two different kinds of functional representational system, one of which is more appropriate for determinate information and the other for indeterminate information. They presented their subjects with a description of a spatial layout of some objects. Some descriptions were determinate, i.e. they described a unique spatial layout, while others were indeterminate i.e. they were true descriptions of several different spatial layouts. Mani and Johnson-Laird found that subjects tended to remember the 'gist' of determinate descriptions better than indeterminate descriptions suggesting that they had been able to construct a 'model' of the determinate layout. The fact that subjects remembered the actual wording of the indeterminate descriptions better than the determinate descriptions suggested that when the subjects were unable to construct a unique model of the layout they tended to retain the verbally presented information but when they could construct a model retention of the verbal description was not so important.

Mental Models

Although it has been claimed that the debate about representation is not resolvable (Anderson 1976, 1978), Johnson-Laird (1980, 1983) has recently described the problem in a different light. He argues that at one level of description it is undoubtedly true that "everything can be reduced to a uniform code in the language of the brain" (Johson-Laird, 1983, p. 165) but from a psychological point of view it is useful to maintain a higher level distinction between functionally different modes of representation. Johnson-Laird considers the analogy of different kinds of data structures used in writing computer programs (i.e. different representational formats) which can nevertheless be reduced to a unitary language of machine code instructions. So too, argues Johnson-Laird, different functional representations in the brain can be reduced to a unitary language - the language of the brain. The distinction between different types of functional representation can and should be maintained at the level of psychological explanation.

Johnson-Laird claims that there are three different types of mental representation: propositional representations, which are strings of symbols similar to the surface linguistic forms of sentences;
mental models, which are structural analogues of the world; and images, which are perceptual analogues of models.

Johnson-Laird introduced the notion of mental models into theories of meaning and inference in an attempt to overcome some of the problems associated with propositionally based theories. Johnson-Laird believes that many of the problems with theories of meaning and models of reasoning based on propositional representations arise because these systems are like uninterpreted predicate calculus formulae and are in need of interpretation. Any propositionally based theory of meaning and inference is consequently incomplete since uninterpreted logical systems are merely systems for manipulating strings of symbols. Among the problems which arise for such systems are:

(a) The problem of how these strings of symbols come to have meaning for the subject.

(b) The related issue of the acquisition of logical rules of inference. There is apparently no way of bootstrapping oneself into a purely formal system.

(c) The problem of errors. Since inference is always presumed to be logical, errors are assumed to occur at the stage of interpretation.

(d) The problem of which logic is the logic used in the mind. As we have seen psychologists have generally adopted the standard logics - propositional logic and predicate logic. However it is widely recognised that these logics are inadequate for dealing with many natural language inferences which a complete theory of reasoning would be required to account for.

(e) Another problem for logically based theories of reasoning is the problem of content. Theories of reasoning which propose that inference is carried out by means of logical rules find it very difficult to explain why subjects find it much easier to make inferences with familiar/concrete content.

Johnson-Laird argues that mental models provide interpretations for propositional representations. It is important for Johnson-Laird that, in contrast with the manipulations of meaningless symbols proposed in syntactic based rationalist theories, the operations carried out during reasoning should be meaningful for the subject since otherwise it is not clear how subjects could learn and use these operations to make inferences. Johnson-Laird's intention was to provide a theory of meaning and inference which obviated the problems associated with logically based theories of reasoning but which did not deny that human beings can make logical inferences.

Johnson-Laird accredits Craik (1943) with the original idea that thinking can be regarded as the manipulation of internal representations or models of the world. The particular mental models which Johnson-Laird describes are obviously similar in many ways to the models used in formal semantics to provide a semantic characterization of a formal language. Before discussing the extent of the relationship between mental models and model theoretic semantics a brief description of model theoretic semantics will be given.

In characterizing a formal language the logician specifies its syntax, the formation rules which determine the well-formed expressions of the language, the rules of inference, which allow the derivation of other formulae and the semantics, which provides an interpretation of the formulae determined by the syntax and inference rules. For the logician, unlike the linguist, the syntax is inevitably a basis for semantic generalization. In extending the use of formal systems to natural language analysis the logician assumes that the syntax of natural language will consist of a specification of the well-formed expressions or logical forms of the language.

In model theoretic semantics the well-formed expressions of a language are provided with a semantic interpretation with respect to a model structure and an assigment function which maps the well-formed expressions defined by the syntax into elements in the model structure. The model structure and assignment function together constitute the model. The ordered pair of a domain of individuals, D, and a valuation function, V, for example, is a model, D,V, for a predicate logical system. The domain is the set of objects which are denoted by individual variables and constants in the formal language. In other words the domain consists of a set of objects which provide referents in the model for terms in the formal language. The assignment function assigns values to the individual terms, predicates and well-formed sentences of the language. In interpreting an individual term, a, of the language, some element, d, of the domain, D, is assigned by the valuation function, V, as the value of that term: V(a)=d. One place predicates are interpreted as sets of objects having certain properties and, more generally, n-place predicates are interpreted as sets of ordered n-tuples (sets with n members) standing in a certain relation to each other. For a one-place predicate for example the valuation function V assigns a set of elements of the domain D_i (where $D_i \subseteq D$) to the predicate f: $V(f)=D_i$. D_i is therefore the set of elements in the domain with the property f. The interpretation of the whole proposition is determined from the interpretation of its components e.g.: $V(f(a))=1$ iff $V(a) \in V(f)$, $V(f(a))=0$ iff $V(a) \notin V(f)$. This says that the proposition f(a) is true if and only if the element in the domain denoted by a (i.e. d_i) is a member of the set of objects in the domain denoted by f (i.e. D_i). A formula that is true in at least one model is called satisfiable while a formula that is true in all models is called a valid formula. Logicians are usually interested in establishing the valid formulae of a formal language.

The model outlined above would be inadequate for dealing with the semantics of natural language. Many richer formal languages have been developed which can deal with a variety of natural language concepts including modal expressions, necessity/ possibility (alethic logic), belief (doxastic logic), obligation and permission (deontic logic) knowledge (epistemic logic) and temporal concepts (tense logic). The model structures required to characterise the semantics of these systems are correspondingly more complex than the semantics for predicate logic. The semantics for modal logic has introduced the notion of a possible world to deal with the interpretation of possibility and necessity. Model structures for possible world semantics thus require in addition to the domain D, a specification of such entities as W, the set of possible worlds; W_i a specific element of W; R, a relation of accessibility defined over members of W. Recent work in the application of model theoretic semantics to natural language, which was initiated by Montague in the 1960s, has been remarkably promising in characterising the semantic structure of many aspects of natural language which had previously not been treated formally.

Despite the success of model theoretic semantics in characterising semantic structures in formal systems and in natural language, the relationship between formal semantics and psychological theories of semantics is very unclear. Lyons (1981) says that it is perfectly legitimate to talk about formal semantics "as if it is the aim of formal semantics to construct models of the mental representations that human beings have of the external world" (Lyons, 1981, p. 165) but acknowledges that this is not how logicians and philosophers would regard formal semantics. It is tempting to suppose that the relationship between formal and psychological semantics should be viewed in light of Peters' "ecumenical principle" which holds that formal semantics characterizes what is computed in understanding a sentence, while psychological semantics specifes how it is computed. Partee (1979) and Johnson-Laird (1982) have dismissed this simple relationship between formal and psychological semantics. The main problem with the view that models in the mind are the psychological couterparts of the models postulated in model theoretic semantics is that the mind could not represent the infinitely many possible worlds proposed in model theoretic semantics. Johnson-Laird's solution to this problem was to propose that "a mental model is a single representative sample from the set of models satisfying the assertion" (Johnson-Laird, 1983, p.264).

Mental model theory is the psychological analogue of formal semantics in model theoretic semantics. In both theories sentences are interpreted by mapping strings of symbols (propositional representations) into models. The mental model correlate of the model structure of predicate calculus would presumably be the restricted set of prototypes which act as referents in the model for terms in natural language. The correlate of the assignment function would presumably be the mental model procedures mapping the terms in the language onto their referents in the model. The procedures in the mental model which set up identities between prototypical elements in the model correspond to the rules for building up interpretations of sentences.

It is useful at this point to look at an example of how mental models operate. Johnson-Laird (1980, 1982) provides a convincing demonstration of the predictive utility of the mental model theory of inference in explaining certain effects found in syllogistic reasoning tasks. Johnson-Laird and Steedman (1978) described the figural effects, previously unreported biases in syllogistic inference associated wth the figure of the premises. Logicians traditionally recognize four figures of the syllogism:

(1)	(2)	(3)	(4)
B-A	A-B	B-A	A-B
C-B	C-B	B-C	B-C
C-A	C-A	C-A	C-A

Johnson-Laird and Steedman point out that the C-A form of the conclusion is arbitrary and in a psychological study conclusions of the form A-C should also be judged as valid. Johnson-Laird and Steedman found that for syllogism forms (1) and (4) there was a bias to drawing conclusions of a particular form: the preferred conclusion for figure (1) was C-A while for figure (4) it was A-C. For syllogism forms (2) and (3) there was no preferred conclusion. Johnson-Laird and Steedman explained this effect in terms of the asymmetry of the connection between the terms in the premises.

In setting up a mental model of a syllogism such as:

(a) "Some of the scientists are parents.

All of the parents are drivers."

the subject will represent the first premise by imagining an arbitrary number of prototypical members of the set of scientists and establishing the identitiy of these with an arbitrary number of prototypical members of the set of parents (Figure 1.3a). The representation of the identity scientist=parent, or generally a=b, is asymmetrical in that it is easier to establish 'parent' ('b') given 'scientist' ('a') than to establish 'scientist' ('a') given 'parent' ('b') because of the characteristics of working memory. The elements in parentheses indicate that the premise "Some scientists are parents." allows for the possibility of some scientists who are not parents and some parents who are not scientists. The information from the second premise is simply added on to the mental model already established (Figure 1.3b). In this syllogism the conclusion "Some scientists are drivers." will be preferred to "Some drivers are scientists.", which is also valid, because the asymmetry in mapping the elements makes it easier to read the model in the direction of the mapping than against it.

Johnson-Laird also predicts an order of difficulty for the different syllogistic figures based upon the number of operations which would be required in order to form an integrated model of the premises. These operations include renewing interpretations of the first or second

Figure 1.3a

scientist = parent

scientist = parent

(scientist)(parent)

Figure 1.3b

scientist = parent

scientist = parent

(scientist) (parent = driver)

(driver)

Figure 1.4

a = b	becomes	b = a
a = b		b = a
(b)		(b)

premises and "switching round" operations which bring the two middle terms together. In syllogistic figure (1) the first premise would be renewed or re-encoded after the second premise in order to bring the middle terms together: thus B-A would be re-encoded after C-B:

$$\begin{matrix} B\text{-}A \\ C\text{-}B \end{matrix} \longrightarrow \begin{matrix} C\text{-}B \\ B\text{-}A \end{matrix}$$

This renewal interpretation makes it clear why the preferred conclusion to this figure is C-A.

In order to bring occurrences of the middle terms together in figures (2) and (3) the switching operation would be required. Thus in figure (2), C-B would be switched to B-C. This switching operation is similar to a conversion operation but differs from it in important respects. The premise "All A are B" for instance would not be interpreted as its converse "All B are A". The switching operation preserves the logical accuracy in that possible b elements are still represented but the direction of mapping is changed from a-b to b-a (see Figure 1.4).

The number of operations required to form an integrated model of the premises for a particular figure was zero for figure (4), one for figure (1), one or two for figure (2) depending on whether the conclusion was A-C or C-A and two or three for figure (3) depending on whether the conclusion was C-A or A-C. Johnson-Laird and Bara (1982) found that correct response on the different figures was indeed a function of the proposed operations with 51%, 48%, 35% and 22% correct on figures (4), (1), (2) and (3) respectively.

Another factor which was found to influence the difficulty of a particular syllogism was the number of mental models of the premises that it is possible to construct. Syllogism (a) above for example is relatively easy to solve because there is only one model of the premises which can be constructed. A syllogism of the form :

b) "All the beekeepers are artists.

None of the chemists are beekeepers."

on the other hand will only yield a valid conclusion if three separate models are considered. The initial model of the premises will look like Figure 1.4a where the first premise is re-encoded after the second in order to bring the middle terms together. Consideration of this model would lead to the conclusion "None of the chemists are artists" and indeed 60% of the responses by Johnson-Laird and Steedman's subjects were of this type. The bias towards accepting C-A conclusions explains why 60% of responses were of this type while only 10% of responses were "None of the artists are chemists". There are other models which can be constructed which satisfy the premises however, for example the model in Figure 1.4b. This simple manipulation of the model shows the falsity of the previous conclusion. The conclusion that follows from this model is "Some chemists are artists". The fact that no subjects made this response while 10% of responses were "Some chemists are not artists" was understood to indicate that the model in Figure 1.4a was constructed prior to the model Figure 1.4b and that the latter conclusion was abstracted from both models. Figure 1.4c shows the third possible model which is compatible with the premises. Subjects who have constructed all three models might respond that there is no valid conclusion i.e. no conclusion common to all three models and indeed 20% of responses were of this type. There is an fact a valid conclusion which is true in all three models, "Some artists are not chemists.", but few subjects made this response indicating that this

<u>Figure 1.4a</u>

chemist
chemist
———————————
beekeeper = artist
beekeeper = artist
(artist)

<u>Figure 1.4b</u>

chemist
chemist = artist
———————————
beekeeper = artist
beekeeper = artist
(artist)

<u>Figure 1.4c</u>

chemist = artist
chemist = artist
———————————
beekeeper = artist
beekeeper = artist

inference is beyond the great majority of subjects.

For those syllogisms which have a valid conclusion, either one, two or three models are required to generate the conclusion. Johnson-Laird and Steedman found that the percentage of valid conclusions drawn was a function of the number of models required. Averaging over three experiments Johnson-Laird reported 78%, 29% and 13% correct responses on problems requiring one, two and three models respectively.

Johnson-Laird argues that the mental model theory of syllogistic inference is superior to rival theories like Erickson's (1974), which was based on Euler circles, or Guyote and Sternberg's (1978), which was based on symbolic representations. According to Johnson-Laird mental models are more plausible as psychological models since they are "natural" models of discourse in contrast with the models of Erickson and Sternberg which are derived from sophisticated mathematical notations for dealing with infinite sets and are not intuitive psychological models.

Johnson-Laird claims that reasoning in terms of mental models is reasoning without logic. Reasoning with mental models allows valid inferences to be made without needing to postulate an internalized set of inference rules. The mental model theory of inference is a semantically based theory of inference which proposes that valid inferences are made by constructing a model as an interpretation of the premises, seeking alternative models which are consistent with the premises and abstracting from these models a conclusion which is true in all models.

The mental model theory of inference provides an explanation of the errors which subjects make in reasoning and also specifies the sequence of operations which would be required to make a valid inference. Inferential ability is held to consist of three components 1) the ability to construct an integrated model of the premises 2) the ability to search for alternative models which are compatible with the premises and 3) the ability to abstract from all models compatible with the premises a conclusion which is true in all of them. Errors in syllogistic reasoning were explained in terms of the subject's failure to consider all alternative models which are compatible with the premises or in terms of their inability to abstract from the alternative models of the premise a conclusion which was true in all models. Presumably there are also errors which arise from the subject constructing an inappropriate model of the premises initially, although this is more difficult to assess.

Johnson-Laird considers that his mental model theory overcomes many of the problems which occur for theories of reasoning which advocate a mental logic. The theory is intended as a general theory of inference which accounts for inferential ability with arguments from standard logical systems, such as the traditional syllogisms, propositional inferences and the quantificational calculus. However it is also the aim of the theory to account for inferences in natural language which are psychologically acceptable but which traditional logical systems have failed to account for. Although this type of inference has traditionally been regarded as non-logical or pragmatic Johnson-Laird does not consider the logical/pragmatic distinction useful psychologically but, as we have seen, distinguishes between explicit and implicit inferences. The valid deductive inferences which Johnson-Laird calls explicit inferences do not just include standard logical inferences but also natural language inferences which are not regarded as logically valid within any standard logical system. The major distinction which Johnson-Laird draws between explicit and implicit inferences is in terms of whether there is an active search for alternative models of the premises or discourse.

Johnson-Laird views the average reasoner in the same way as he views the early logician attempting to determine the valid inferences of a particular logical system. In attempting to discover whether a particular inference is valid the reasoner would try out different terms in an argument while keeping the form of the argument constant e.g. he might argue:

"All horses are animals.
Some animals are browm.
Therefore some horses are brown."

While this is a true statement it is not a valid conclusion, as the reasoner will discover if he attempts to substitute the terms of the argument e.g.

"All horses are animals.
Some animals are carnivores.
Therefore some horses are carnivores."

Since a false conclusion cannot follow from true premises this particular form of argument is rejected as invalid. Having established a valid set of arguments by these essentially semantic methods the logician can then abstract and formulate a set of valid inference rules. Johnson-Laird claims that the reasoner constructs valid inferences in a similar way by evaluating the truth or falsity of the conclusion to a particular argument form with different kinds of content. Obviously the average reasoner is not as far advanced as the logician in abstracting valid principles of reasoning from repeated exposure to instances of particular arguments. Cohen (1981) argues in a similar way that the normative criteria by which the validity of inferences are evaluated have been derived from a formalisation of intuitions about validity.

The developmental aspect of the mental model theory is largely unexplored. Johnson-Laird believes that mental model theory provides a more plausible explanation of how children acquire the ability to make inferences than theories which advocate a mental logic such as Piaget's or theories which propose that logical competence is innate (Fodor, 1980). According to Johnson-Laird children cannot learn to make valid inferences until they have learned the truth conditions of linguistic expressions. Presumably the development of inferential ability would be explained by the increasing ability to construct models of the premises which are increasingly accurate representations of the truth conditions of the premises. i.e. by an improvement in the first component of the three component reasoning skills described above. Improvement in the other two abilities is more likely to occur after the subject has become reasonably proficient at constructing models of the premises in the first place.

Johnson-Laird acknowledges that many people will assume that logical inference rules are somehow embedded in the theory of mental models but he is adamant that this assumption is untenable. Falmagne (1975) suggested that in models of logical reasoning in which information from the premises is formalized into concrete models, such as Erickson's (1974) Euler diagram theory, Huttenlocher's (1968) spatial imagery theory, Johnson-Laird's (1981) mental model theory, and in which logic appears not to be involved in the traditional sense, that logic is not absent but is hidden in the formalization stage. In a strictly logically based model of reasoning the natural language input is formalized, i.e. the logical form of the premises is abstracted; logical inference rules can then be applied to the well-formed expressions to derive valid conclusions. When information from the premises is formalized into a spatial model there is a sense in which the conclusion can be simply "read off" from the model rather than derived from the inference rules in the standard way. Falmagne argues that in such models the locus of logic has been shifted from the system of inference rules to the formalizaton process.

In choosing the appropriate spatial model to represent the premises, e.g. a straight line in a transitive inference task or an appropriate mental model in a syllogism task, the subject demonstrates his knowledge of the formal similarity between the expressions of the language and the relevant logical properties of the model. This knowledge is also implicated in reading off the conclusion from the model. In claiming that the logical burden in these concrete models lies in the formalization process Falmagne means that the logical properties of expressions which contribute to the determination of the set of permissible inferences are inherent in the representation, For example, in choosing Euler diagram representation 5a rather than 5b or mental model representation 6a rather than 6b as an interpretation of the universally quantified statement "All A are B", the subject is determining that some inferences will be permissible while others will not be. If the second premise is "No A are C", for instance, models 5b and 6b will determine that any relationship between B and C such that "Some B are C" or "Some C are B" is disallowed while models 5a and 5b would permit such a relationship.

Falmagne regards the subject's choice of representation, be it a

Euler circle representations of "All A are B."

Figure 1.5a

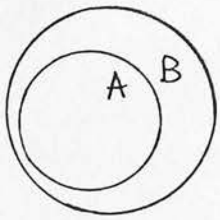
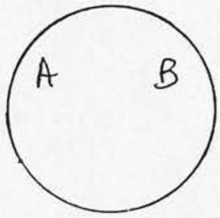


Figure 1.5b



Mental Model representations of "All A are B."

Figure 1.6a

a = b

a = b

(b)

"a" denotes a prototypical exemplar of a member of set A.
"b" denotes a prototypical exemplar of a member of set B.
"=" denotes an identity.
parentheses indicate that the relevant individual may not exist.

Figure 1.6b

a = b

a = b

logical formalism or a concrete model, as reflecting his knowledge of the formal similarity between the premises and that representation. It is important to emphasise that Falmagne defines logical competence as including not just competence with logical relationships and inference rules but also a competence in translating the verbal problem into the appropriate representational format.

Johnson-Laird (1980) claims that the logical component of the mental model theory "consists solely in a procedure for testing models". In testing mental models alternative models are constructed which are compatible with the premises and an exhaustive search for such models would yield valid inferences. By 1982, however, he seems to want to include the procedures for setting up mental models as well as the procedures for testing them as the locus of the logical component: "The logical properties of an expression are not directly represented in the mind but emerge naturally as a consequence of the use of the expression in the construction and search procedures." (Johnson-Laird, 1982, p. 12).

For Falmagne logical knowledge is inherent in the subject's choice of representation. Johnson-Laird on the other hand regards logical knowledge as an emergent property of the construction and manipulation of mental models. For Falmagne the subject appears to have discovered the logical properties of language and can use this knowledge to make valid deductions while for Johnson-Laird the subject has to construct a model of the linguistic premises (an interpretation) from which the logical properties will emerge, but only if exhaustive tests are carried out. Falmagne would argue that if logical properties were not inherent in mental models then mental models would be an ineffective medium for generating valid deductions. These two positions will be discussed further later in the thesis.

PART 2: THE ACQUISITION OF DEDUCTIVE REASONING ABILITY

INTRODUCTION

So far very little has been said about how the ability to make deductive inferences might be acquired. Obviously any account of the acquisition of this ability should examine the nature of the ability: this was the intention of Part 1 of this chapter. However little concensus of opinion was found in the literature concerning how people make deductive inferences, with some arguing that human reasoning is inherently rational (Henle, 1962; Revlis, 1975) and that thought processes can be identified with logical inference rules or natural deduction schemes (Osherson, 1974; Braine, 1978) and others arguing that non-logical response biases (Evans, 1978, 1980) based on pragmatic response or availabilty effects (Manktelow and Evans, 1979; Pollard, 1981) provide better accounts of performance in reasoning tasks. A weaker, more plausible version of the rationalist hypothesis which has recently emerged is that subjects have an underlying logical competence which can be characterised in terms of a logical system but a variety of performance factors influence actual responses made in reasoning (Cohen, 1981). The most recent and controversial theory is Johnson-Laird's (1980, 1983) mental model theory which proposes that the ability to make deductive inferences should be explained in terms of the ability to construct and manipulate mental models of the premises.

Different accounts of the nature of deductive reasoning ability will obviously require different accounts of how this ability is acquired. Rationalist models of acquisition have tended to argue that logical principles are innate (Fodor, 1980) or are acquired in the same kind of way as the principles of grammar (Falmagne, 1980). Both accounts are difficult to test. Although Johnson-Laird (1980) claimed that his mental model theory obviates many of the problems inherent in rationalist accounts of reasoning and consequently provides a better account of the acquisition of the ability to make deductive inferences than rationalist accounts, the developmental aspect of his theory has not yet been explored. Piaget's account of the development of logical thinking has dominated the literature on the development of children's thinking and will be discussed at length since other developmental theories are inevitably compared and contrasted with Piaget's. Piaget's theory is essentially a rationalist theory and therefore susceptible to the general criticisms of rationalist theories but recent interpretations of Piaget's theory have proposed that it should be regarded as a theory of logical competence (Brown and Desforges, 1978; Wildman and Fletcher, 1979; O'Brien and Overton, 1982).

Falmagne (1975) points out that the relative dearth of research on the development of deductive inferential ability with self-contained verbal premises is largely attributable to the strong influence of Piaget's conviction that children under the age of 11 to 12 years cannot handle propositional reasoning problems. The research that there was discussed children's responses on propositional reasoning tasks in terms of whether they provided evidence for or against Piaget's theory. More recently some researchers have questioned the relevance of Piaget's theory to explaining performance on purely verbal inference tasks (Ennis, 1975; Wason, 1977; Braine, 1978; Falmagne, 1980).

Piaget's Theory of Logical Development

Piaget's exposition of his theory of developmental stages, including descriptions of the structural characteristics of each stage and the transition from one stage to the next, is very complex and has often been misrepresented and misunderstood. It is the aim of this thesis to clarify and evaluate Piaget's claims concerning the abilities of children at the concrete and formal operational stages to comprehend and reason with universally quantified and conditional statements in the light of his claims about the cognitive structures at these stages.

Piaget regarded himself as a genetic epistemologist which can broadly be defined as one who studies the development of knowledge both in the sense of scientific bodies of knowledge and in the sense of the development of an individual's knowledge. Piaget proposed that the child acquires knowledge by acting on his environment and, by observing the results of these actions on the environment, adapting these actions. The child is seen not as passively assimilating knowledge but rather as constructing his knowledge of reality. Piaget proposed four qualitively distinct stages which occur in the course of cognitive development. Each stage has a different structure and the order of the stages is invariant since the structure of each stage is constructed from and therefore dependent upon that of the previous stage.

The first stage, the sensory motor stage, lasts from birth to 18 months to 2 years. During this time the infant gradually modifies his simple reflex capabilites (sucking, grasping and swallowing) and, through experience, expands and combines these modified actions to produce new forms of behaviour. By the end of this period the child has constructed a "logic of action" the practical precursor of his later logical operational thinking. The end of this stage is marked by the ability to think symbolically i.e. to internally represent actions and events without being dependent on the actual physical presence of these actions and events.

The pre-operational stage, from 2 to 7 years, spans the interval from the advent of symbolic representation to the advent of operational thinking. The operational stage is marked by the organization of thoughts into interrelated systems but the thinking of the pre-operational subject lacks this systematization and consequently, although in some aspects the child's thinking seems rational, his thinking is full of inconsistencies, he is easily misled, and he is egocentric.

The concrete operational stage (7 to 11-12 years) marks the beginning of logical operational thinking. By operations Piaget means internalized reversible actions i.e. actions which have been represented as mental processes. These operations can only be sensibly understood as operations within an integrated system of interrelated acts. The concept of classification operations for example cannot be interpreted as isolated entities but are only meaningful within a classification system.

At this stage the child's thinking appears to reach a relatively stable equilibrium which Piaget describes in terms of logicomathematical structures. Piaget believes that the cognitive structures at this stage of operational thinking are most accurately modelled by a logical analogue of the mathematical group which he calls a grouping (a hybrid structure combining properties of mathematical groups and lattices). Piaget describes nine groupings including a preliminary grouping of equalities, four groupings which deal with the addition and multiplication of classes and four corresponding groupings dealing with addition and multiplication of relations. Of particular relevance to this thesis are: grouping 1, the additive composition of classes; grouping 2, the secondary addition of classes and grouping 3, the bi-univocal multiplication of classes. These are described below.

Grouping 1, the additive composition of classes describes the operations involved in understanding hierarchical classification, for example

labradors (A) + all dogs other than labradors (A') = dogs (B)

dogs (B) + all carnivores other than dogs (B') = carnivores (C)

carnivores (C) + all animals other than carnivores (C') = animals (D)

Mastery of this grouping enables the child to perform successfully on a variety of hierarchical classification tasks; the cognitive operations involved in these tasks can be symbolized by the logical addition and subtraction of classes e.g. the symbolism:

$(+A) + (A') = (+B)$

corresponds to the mental operation involved in thinking of a subclass, e.g. labradors, and its complement, the set of dogs that are not labradors, and realizing that this must be equivalent to the including class, dogs.

Grouping 2, the secondary addition of classes is concerned with the alternative division of a class into subclasses or what Piaget calls vicariances. The class of dogs (B) for example can be divided into subclasses: spaniels (A_1) and non-spaniels (A_1'); labradors (A_2) and non-labradors (A_2'); poodles (A_3) and non-poodles (A_3') etc. where $(A_1 + A_1') = (A_2 + A_2') = (A_3 + A_3') = B$

Grouping 3, the bi-univocal multiplication of classes is concerned with the division of a class into subclasses along two different dimensions and the multiplication of these subclasses. For instance a class of people can be divided in terms of skin colour: A_1 = white, B_1 = black, C_1 = yellow, where $A_1 + B_1 + C_1 = D$. The same class of people can be subdivided along a different dimension e.g. where they live A_2 = urban, B_2= suburban, C_2= rural where $A_2 + B_2 + C_2 = D$. $(A_1 + B_1 + C_1) = D = (A_2 + B_2 + C_2) = D$. (This is a vicariance). Logical multiplication generates the following products or "base associations" as Piaget calls them: $(A_1+B_1+C_1)(A_2+B_2+C_2) = A_1A_2 + A_1B_2 + A_1C_2 + B_1A_2 + B_1B_2 + B_1C_2 + C_1A_2 + C_1B_2 + C_1C_2$.

Inhelder and Piaget (1964, p.151) acknowledge that logical multiplication has a more complex logical structure than logical addition but claim nevertheless that "multiplicative and additive classification are mastered at the same time - about the age of 7 - 8 years". The rationale for this claim is that the graphic properties of multiplicative classification compensate for their increased complexity compared with additive classification.

Piaget claims that the logicomathematical structures he calls groupings model the cognitive structures and operations of the concrete child and he provides detailed descriptions of the groupings. He has often been criticised for providing an analysis which concentrates too much on logical structures and not enough on psychological factors which determine performance on tasks requiring concrete operational thinking (Bryant and Trabasso, 1971; Boden, 1979; Donaldson, 1978). Even some of his followers have questioned the existence of the operational structures (Smedsland, 1977) and their utility in explaining behaviour in tasks which are supposedly related by having the same underlying structure. Although there has been a lot of research concerned with some of operational structures, e.g. the additive composition of classes and the addition of symmetrical relations, it is difficult to translate some of the operational structures, e.g.the bi-univocal multiplication of relations, into empirical tasks. For those tasks which have been extensively studied there is evidence that performance on tasks which can be described in terms of the same logical structure is different for problems with different content. This is incompatible with Piaget's proposals that the logical structure, not the content, of a task determines performance. The concept of horizontal decalage has been introduced into Piaget's theory to cope with the problem of different levels of performance on problems with the same logical structure but different content.

Piaget has also been accused of logicism (Ennis, 1975; Brainerd, 1976-77) but he did not consider logic and psychology to be co-extensive as is evident from his conclusion in Mathematical Epistemology and Psychology that "logic, as a theory of demonstrative reasoning, is completely autonomous in relation to psychology, and vice versa." (p.305). However he frequently talks as if the groupings he describes correspond to mental operations carried out in reasoning. Formal operations, like concrete operations, can be described in terms of logicomathematical structures. Just as the cognitive structures of the concrete operational child can be modelled by a set of "groupings" so the cognitive structures of the formal operational individual are modelled by a fully integrated lattice-group structure - the "structure d'ensemble". This structure is derived by an extension of the classification operations of concrete operations to include a classification of the products of concrete operational classification i.e. operations upon operations. This classification of classifications is what Piaget refers to when he describes formal operations as "second degree or second order operations".

The concrete operational child can logically multiply two classes (A +

A') and (B + B') to obtain the products AB + AB' + A'B + A'B', the subclasses which Inhelder and Piaget (1958) call the "base associations"; this will be recognized as involving grouping 3, the logical multiplication of classes. The formal operational subject can then take these base associations as elements in a further classification operation or what is called a "combinatorial analysis". That is he can combine them 0, 1, 2, 3 or 4 at a time to give all possible arrangements of base associations. This n*n classification of base associations will be recognized as involving grouping 2, the division of a class into subclasses. There are in fact 16 possible combinations of four elements. Whereas concrete operations were tied to concrete objects and events the combinatorial system enables the subject to combine objects, factors or propositions; the combination of propositions gives rise to a new logic of propositions which enables the child to think hypothetico-deductively. When the elements of the combinatorial analysis are propositions this analysis will yield the sixteen binary combinations of propositional logic. Each combination of elements corresponds to a particular operation of propositional logic. The combinations and the propositional operations to which they correspond are shown in Table 1.1. The combination (p.q) for example corresponds to the propositional operation of conjunction; The combination (p.q) (-p.q) (-p.-q) corresponds to the propositional operation of implication $p \supset q$. The sixteen possible propositional operations form a lattice structure with the combinations themselves as elements. This lattice structure is shown in Table 1.2.

A further advance of the formal operational thinker over the concrete operational thinker is in the integration of the two separate forms of reversibility present at concrete operations, inversion and reciprocity, into a single more powerful system - the INRC group. The propositional combinations are all inter-related in terms of the

TABLE 1.1

PIAGET'S SYSTEM OF SIXTEEN BINARY OPERATIONS

CONSTRUCTED COMBINATION	NAME OF OPERATION
1 NEGATION OF COMPLETE AFFIRMATION	-
2 CONJUNCTION	(p.q)
3 NONIMPLICATION	(p.-q)
4 NEGATION OF RECIPROCAL IMPLICATION	(-p.q)
5 CONJUNCTION NEGATION	(-p.-q)
6 AFFIRMATION OF p	(p.q) (p.-q)
7 AFFIRMATION OF q	(p.q) (-p.q)
8 EQUIVALENCE	(p.q) (-p.-q)
9 RECIPROCAL EXCLUSION	(p.-q) (-p.q)
10 NEGATION OF p	(-p.q) (-p.-q)
11 NEGATION OF q	(p.-q) (-p.-q)
12 IMPLICATION	(p.q) (-p.q) (-p.-q)
13 RECIPROCAL IMPLICATION	(p.q) (p.-q) (-p.-q)
14 DISJUNCTION	(p.q) (-p.q) (p.-q)
15 INCOMPATIBILITY	(p.-q) (-pq) (-p.-q)
16 COMPLETE AFFIRMATION	(p.q) (-p.q) (p.-q) (-p.-q)

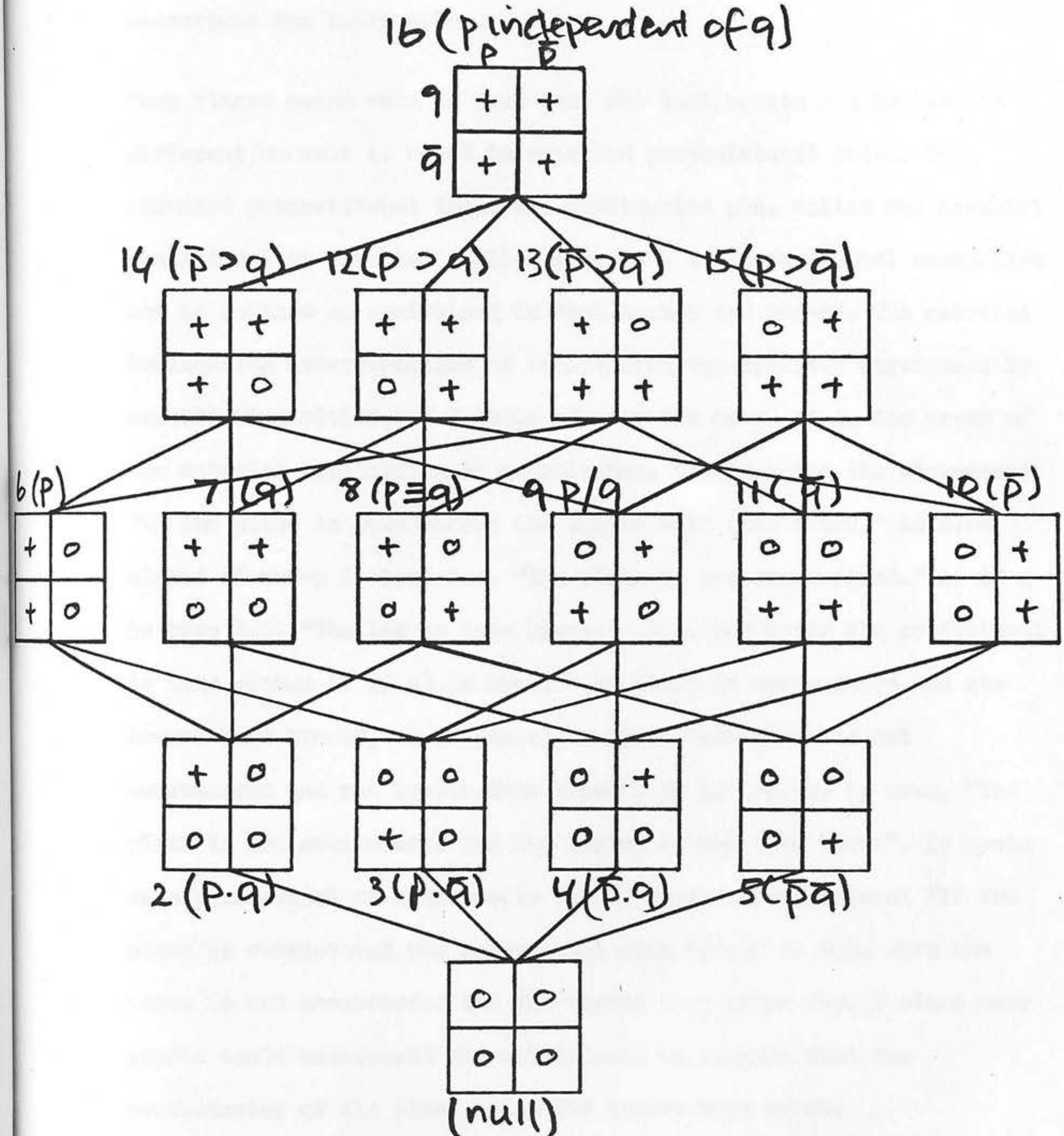


Table 1.2

The lattice structure of the truth tables for the 16 binary operations

Numbers refer to operations in Table 1.1

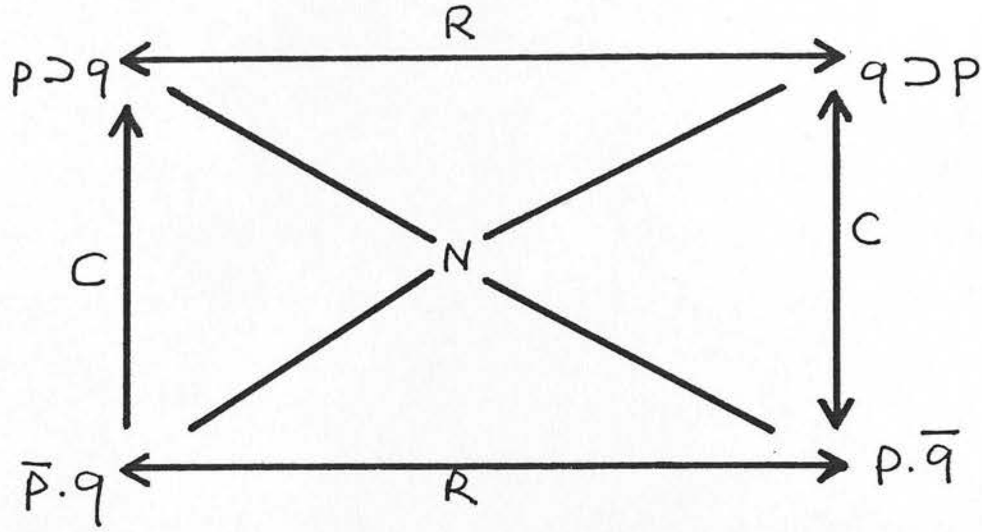
identity, negation, reciprocity and correlative transformations. Figure 1.7 shows the group structure of the relationships among the operations for implication.

What Piaget means when he says that the implication p⊃q is true is different from what is meant in standard propositional logic. In standard propositional logic the implication p⊃q, called the material conditional or material implication, is a truth functional connective and is defined as equivalent to "not both p and not-q". The material implication interpretation of implication incorporates paradoxes. By establishing either the falsity of p or the truth of q, the truth of the material implication is established. For instance the statement: "If the plant is overwatered the leaves will turn brown." is true either if not-p is true i.e. "The plant is not overwatered." or if q is true i.e. "The leaves turn brown." In other words the conditional is true either if (p.q) is true, "the plant is overwatered and the leaves turn brown", or if (-p.q) is true, "the plant is not overwatered and the leaves turn brown", or if (-p.-q) is true, "the plant is not overwatered and the leaves do not turn brown". It would seem paradoxical to many people to say that the conditional "If the plant is overwatered the leaves will turn brown" is true when the plant is not overwatered and the leaves turn brown (-p.q) since many people would understand the conditional to suggest that the overwatering of the plant makes the leaves turn brown.

In an attempt to understand Piaget's use of the symbolism of propositional logic Ennis (1975, 1976) suggests that it was perhaps Piaget's intention in interpreting propositional logic in the way he does to avoid the paradoxes of material implication. For Piaget to show that a combination, e. g. (p.q) (-p.q) (-p.-q), is true it must be shown that instances of all elements in the combination, e.g.

Figure 1.7

Group Structure for relationships among operations for implication



(p.q), (-p.q) and (-p.-q), "exist" while no instances of elements not in the combination, e.g. (p.-q), exist. Obviously Piaget does not interpret p and q in the standard way as propositions since if (p.q) is shown to be true (-p.q) cannot also be true at the same time. It cannot both be the case that (p.q), the plant is overwatered and the leaves turn brown, and (-p.q), the plant is not overwatered and the leaves turn brown. Parsons (1960) claimed that Piaget's system does not make logical sense. He points out that Piaget uses the letters p and q to represent propositional functions or open sentences rather than propositions, so that rather than p representing the proposition "The plant is overwatered." where "the plant" has a definite referent Piaget uses p to represent a propositional function or open sentence "plant x is overwatered". The truth or falsity of an open sentence depends on the value assigned to the variable x and consequently the truth value can change depending on the situation.

Although he uses the symbolism of propositional logic, Piaget's logic is not truth-functional and, although he does not acknowledge the fact, Piaget uses the device of existential quantification to close his open sentences thereby making them interpretable. The open sentence Px, "plant x is overwatered" will become xPx "A plant is overwatered" or "There is a plant which is overwatered." This is what was meant when it was said that for a Piagetian combination to be true the elements of the combination had to "exist". Parsons points out that the propositional formula p>q and its complete disjunctive normal form p.q -p.q -p.-q express the same logical operation but Piaget tried to establish a correspondence between (x)(Px Qx) and the formula: (x)(Px.Qx).(x)(-Px.Qx).(x)(-Px.-Qx).-(x)(Px.-Qx). Parsons indicates that these formulae are not equivalent since the former states that for all x, if x is a p then x is a q, while the latter states that there is an x which is p and q, there is an x which is not-p and q, there is an x which is not-p and not-q but there is no x which is p and not-q. The formula (x) (Px Qx) does not make existential claims in the way Piaget proposes and for instance could be true even if there were no p.q elements.

Although the interpretation discussed above is a very strict interpretation of Piaget's account of the formal operational combinations it is the interpretation which Piaget's interpreters (Flavell, 1963; Papert, 1963; Ginsberg and Opper, 1969) and critics (Parsons, 1960; Ennis, 1975, 1976) have come up with. But would a formal operational subject really have to establish that, for instance, a thin flexible rod (p.q), a thick flexible rod (-p.q) and a thick inflexible rod (-p.-q) exist and a thin inflexible (p.-q) does not exist in order to say that an implication sentence like "If a rod is thin then it is flexible" is true? It is important to Piaget that he would because by establishing existence and non-existence of elements the formal operational subject establishes the truth of a certain combination and distinguishes it from all other possible combinations. For instance the formal operational subject would be able to discover whether an implication or equivalence relationship existed between two propositions p and q by establishing whether the combination (p.q) (p.-q) (-p.-q) or the combination (p.q)(-p.-q) was true. The crucial difference between an equivalence and an implication is whether the element (-p.q) exists or not. The Piagetian combinations specify the truth conditions for different propositional relationships.

Ennis notes that although the above seems to be the accepted interpretation of Piaget's combinations there is a suggestion in The Growth of Logical Thinking from Childhood to Adolescence (1958) that the possibility of existence of the elements in the combinations rather than their actual existence is all that is required for a particular combination to be true. Ennis rejected the possibility of existence interpretation as incorporating paradoxical requirements. Ennis's objections ran along the lines that it was unreasonable to require the possibility of a thick flexible rod in order for it to be true that if a rod is thin it is flexible. This was also a problem for the "existence" interpretation. A major problem with the possibility of existence interpretation which Ennis does not discuss is the problem of how the combinations would ever be learned if elements in the combinations only possibly existed. If specific elements were not found it would be unclear whether they were not allowed, e.g. the (-p.q) element for the equivalence combination, or whether they were possible elements but simply not present on that occasion.

If Piaget's theory is to be useful as a psychological theory of reasoning the logical structures which characterise the different stages should have observable behavioural correlates. It would be tempting to suppose that the cognitive operations typical of the concrete and formal operational stages would be reflected in the subject's language. Piaget's[1] view on the relationship between language and cognition has always been that language is dependent upon cognition. It might be supposed then that the cognitive operations typical of each stage will determine the subject's understanding and use of language. Indeed it was Inhelder and Piaget's original proposal that children at the concrete and formal operational stages would use different linguistic expressions. Specifically children at concrete operations would not be expected to use propositional connectives such as 'if..then' 'either .or' etc. but rather their language would be expected to reflect their understanding of classification and relational operations, e.g "Maggie is tougher than Neil.", "All men are strong.".

Inhelder and Piaget looked at the statements that subjects at concrete and formal operations produced in their attempts to solve some of the standard Piagetian formal operational tasks and they attempted to reduce the statements to the logical formulae of classes and relations or propositions. They found it almost impossible to say for a particular statement whether it involved class and relational operations or propositional operations. As an example they offer the statement: "This rod bends because it is made from steel." and argue that it is not possible to say whether the cognitive operations underlying the production of this statement involve a recognition of the transitivity of class inclusion: "All steel bars are flexible. This bar is steel. This bar is flexible" or whether it involves a true implication: "If a bar is steel it is flexible.". Inhelder and Piaget consequently reject a purely linguistic criterion as an index of a subject's operational level since it is always possible to express propositional operations in the language of classes and vice versa. They claim that although a concrete operational child may produce statements with propositional connectives he has no real understanding of the propositional relationship involved but rather that his statement reflects the simpler operations underlying class inclusion or relational operations e.g. a concrete operational subject saying for instance "If a bar is steel it is flexible" is simply using a propositional connective to express simpler cognitive operations of class inclusion. The formal operational subject producing such a statement is presumably understood to produce this on the basis of a combinatorial analysis of the situation i.e. he will have found flexible steel bars (p.q), flexible, non-steel bars (-p.q) and inflexible non-steel bars (-p.-q) but no inflexible steel bars (p.-q). Inhelder and Piaget also argued that the formal operational subjects' use of class and relational language reflects underlying propositional operations. The different cognitive structures of concrete and formal operational subjects are not reflected in a straightforward way in terms of the subject's production of language: an isolated statement cannot be identified as a "concrete" statement or a "formal" statement.

Inhelder and Piaget's discussion and rejection of the linguistic criterion is framed in terms of language production. Presumably Inhelder and Piaget would also want to dismiss a linguistic criterion for the comprehension of language. They would not predict, as some have claimed (Roberge and Paulus, 1971; Ennis, 1975, 1976) that a concrete operational subject would be able to understand statements in the language of classes and relations but not statements containing propositional connectives. Rather Inhelder and Piaget would claim that a subject's language comprehension, like his production, is limited by the nature of the cognitive operations available to him.

Two other ways of determining a subject's operational level are discussed by Inhelder and Piaget (1958). The first is to compare all the statements and actions of a subject in solving a task. The second involves analysing the proofs offered by the subject in the solution of a task or an examination of the subject's 'train of reasoning'. In practice the two methods do not seem to be different: "comparing all statements and actions of a subject" would seem to be what one does in "analysing the subject's proof".

By comparing all the statements and actions made by a subject, Piaget proposes that it then becomes evident whether the child is limited to recording his empirical observations of classifications and relations, or whether he is able to separate out variables. The ability to isolate and test variables, a basic scientific reasoning skill, is regarded by Piaget as a criterion for determining whether or not a child is working within the combinatorial system. According to Inhelder and Piaget, separating out variables "implies both hypothetico-deductive reasoning and a combinatorial system". The reason for this is rather difficult to understand, but is possibly easier to explain in the context of an example.

Suppose that a baker is trying to develop a fool-proof recipe for a cake. He might experiment by making up several different recipes, as shown in Table 1.3, which either contain two eggs or three eggs, and which either contain or do not contain baking powder. The resulting mixtures either rise or sink in the cooking. From all the 8 possible outcomes of whether the cakes rise or sink when eggs and baking powder are included or excluded only 4 kinds of results were actually observed, as seen in the Table 1.3. The problem is to decide what contributed to the success of a cake (i.e. its rising). The concrete operational child would be able to classify the data as follows: he would be able to see that the following associations occur:

2 eggs / baking powder: cake rises

2 eggs / no baking powder: cake sinks

3 eggs / baking powder: cake rises

3 eggs / no baking powder: cake sinks

The concrete operational child would be unable to go much further than this classification of empirical observations. The advantage that the

Table 1.3

EGGS	B.P.	POSSIBLE OUTCOME	OBSERVED OUTCOME
2	Y	RISES	Y
2	Y	SINKS	N
2	N	RISES	N
2	N	SINKS	Y
3	Y	RISES	Y
3	Y	SINKS	N
3	N	RISES	N
3	N	SINKS	Y

Variables (eggs and baking powder (B.P.)) and outcomes in isolation of variables example.

formal operational child has over the concrete operational child, according to Piaget, is that he has the ability to isolate and test variables and hence to formulate and test hypotheses. Let us suppose that prior to the experiment the subject guessed that baking powder makes no difference to the cakes, but that cakes with 3 eggs will rise. In order to test his hypothesis "All cakes with 3 eggs will rise", the effect of baking powder has to be excluded. In this case the effect of baking powder can be physically excluded by looking only at those cakes which contain no baking powder. By keeping the effects of baking powder constant (in this case, absent) the subject can test the effects of number of eggs on whether the cake rises or sinks. He discovers the following associations: cakes with 2 eggs which rise (p.x), and cakes with 2 eggs which sink (p.-x), cakes with 3 eggs which rise (-p.x), and cakes with 3 eggs which sink (-p.-x). The formal operational subject knows what the combination (p.x V p.-x V -p.x V -p.-x) means, because he "is working within the combinatorial system". He knows that this combination, one of the 16 possible binary combinations of propositional logic, represents complete affirmation (see Table 1.1), that is, number of eggs and rising or sinking are independent factors.

Having made the initial observations of which outcomes are actually observed as a function of the number of eggs and presence or absence of baking powder as shown above, the subject will possibly change his mind at this point and predict that it is baking powder, not 3 eggs, which makes the cakes rise. (Subjects are notoriously set on trying to verify hypotheses, so possibly many would not change their initial hypotheses, even in the face of apparently conflicting data!)

The subject will now hypothesise that "All cakes with baking powder will rise" and will attempt to test this. In this particular case, of course, it is not possible to physically exclude the effects of eggs: the subject must decide instead to hold the factor "eggs" constant at a particular value while testing the effects of baking powder on the rising or sinking of the cakes. This technique of holding factors constant is critical if the subject is to discover how a particular factor influences the outcome independently of other factors. The subject may decide to look at how baking powder influences the rising/sinking of the cakes when 2 eggs are included in the recipe: he would find the associations q.x and -q.-x i.e. when baking powder is added the cakes rise, when baking powder is not added the cakes sink. In this case the same associations would have been found if the factor eggs had been held constant at "3 eggs". The formal operational child working within the combinatorial system knows that the combination (q.x V -q.-x) represents equivalence: when baking powder is added the cake rises; when baking powder is not added the cake sinks. Of course, as we have already seen, the propositional relationship of equivalence is different from the propositional relationship of implication, which would have held if the hypothesis "All cakes with baking powder will rise" had been true. The equivalence relationship corresponds to the combination (q.x V -q.-x) while the implication relationship corresponds to the combination (q.x V -q.-x V -q.x). According to Piaget, the formal operational subject can distinguish one logical combination from another and would conclude on the basis of the experiment that "All cakes with baking powder will rise and all cakes that rise have baking powder", or "All and only those cakes with baking powder will rise". In practice it is possible that the subject who proposes "All cakes with baking powder will rise" will not regard the absence of a -q.x association, a cake with no baking powder which rises, as disconfirming his hypothesis, although according to Piaget, he should, since it is not just the absence of q.-x instances but also the presence of -q.x instances which are important for the interpretation of this combination. Of course it is possible that there would be a more complex effect of baking powder and eggs such that for example only cakes with either 3 eggs or baking powder but not both would rise. The formal operational subject should be able to deal with interactive effects too. It seems likely that interactive effects, being more complex than main effects, should be more difficult to deal with but Piaget does not mention this aspect of hypothesis testing.

Ennis (1975, 1976) interprets Inhelder and Piaget's claim that the attempt to separate out variables implies hypothetico-deductive reasoning and a combinatorial system as meaning that the isolation of variables strategy is a hallmark of formal thinking i.e. the subject who attempts to isolate and test variables demonstrates that he is "working within the combinatorial system" of formal operations since use of this strategy does not make sense unless the subject possesses the "structured whole" within which he can interpret the particular combination of elements that he finds.

Although Piaget seems to argue (Inhelder and Piaget, 1958, p. 279) that the isolation of variables strategy is an index that the subject has acquired the combinatorial system he undermines the presence of this strategy as an index of combinatorial thinking by arguing that the subject who has not yet acquired this system will be led to its imminent discovery by using the isolation of variables strategy: "isolation of variables ....necessarily ends up at a combinatorial system." (Inhelder and Piaget, p. 288). The isolation of variables strategy cannot be regarded both as a criterion of combinatorial thinking and also as a means of acquiring this system.

Ennis (1975, 1976) cites the evidence of Anderson, 1965 that children under 11-12 years can isolate and test variables. However Piaget himself acknowledged that isolation of variables by negation, i.e. by exclusion of the effects of a factor in one case and inclusion in another case, can be found at the concrete level. Isolation of variables by neutralizing, i.e. by holding the effects of one variable constant for the purposes of testing another variable, is not found until the formal operational level.

Ennis argued that the ambiguities associated with Piaget's theoretical account of the relationship between the isolation of variables strategy and the combinatorial system made the isolation of variables strategy inadequate as a criterion of formal operational thinking.

Ennis claims that Inhelder and Piaget put forward another criterion of formal operational thinking - the ability to distinguish one operation from another. According to Piaget the formal operational thinker can understand and differentiate one from another the sixteen binary operations of propositional logic. Inhelder and Piaget illustrate this with the protocol of a subject from one of their reasoning tasks (the Role of Invisible Magnetism) in which they claim to identify examples of the sixteen binary operations of propositional logic. Ennis criticises this approach on the grounds that it is they, Inhelder and Piaget, who are distinguishing one operation from another, not the subject. Inhelder and Piaget would presumably defend their approach by arguing that they are merely making explicit the operations which can be identified from the subject's proof.

Identification of the operations from the protocol is not as clear-cut as Inhelder and Piaget suggest. Bynum, Thomas and Weitz (1972) examined the same protocol and could only distinguish eight of the sixteen operations identified by Inhelder and Piaget. It is difficult to see how a subject's ability to distinguish one operation from another could ever be conclusively demonstrated since what the subject says and does in the solution of a problem is very difficult to interpret unambiguously. How would a subject express the underlying logical combination, (p.q) (p.-q) (-p.q) (-p.-q), corresponding to complete affirmation for example? Conversely would we want to say that a subject who said that "p is independent of but compatible with q" was actually expressing the underlying operation of complete affirmation. In the example above it was claimed that it was unlikely that a subject would require a -p.q example (a cake with 2 eggs which rises) to be present in order to say that "All cakes with 3 eggs rise.". Subjects simply do not seem to be as precise in distinguishing one operation from another as Piaget's theory requires.

Evidence from many studies showing improvement in performance at adolescence on a variety of 'formal operational' tasks, including combinations (Barratt, 1975; Martorano, 1977; Roberge, 1976; Roberge and Flexer, 1979), isolation and testing of variables (Kuhn, 1977; Kuhn and Angelev, 1976), propositional logic (Kuhn, 1977; Roberge, 1976; Taplin, Staudenmayer and Taddanio, 1974) and proportionality (Brainerd, 1971; Martorano, 1977) supports Piaget's claims about the intellectual advances at this age. There is less agreement however concerning the nature of, or even the existence of, a unifying structure underlying the putative formal operational abilities. Some of the problems which Ennis described in establishing a criterion of formal thinking have been discussed. Others too have found it difficult to generate unambiguous testable predictions from Piaget's theory concerning how to identify behaviour unique to the formal thinker. Roberge and Flexer (1979) point out that although the ability to generate all possible combinations of n elements is used as a standard test of formal thinking it is not clear whether Inhelder and Piaget (1958) intended this or the ability to understand the combinatorial system of sixteen binary operations as an index of formal thought. Lunzer (1965) supported Piaget's view that the distinctive characteristic of formal operational thinking was the recognition of second order relations, i.e. a recognition of relations between relations. By 1978 however he rejected the view that the intellectual advances of the formal operational subject are related by a common underlying structure. He did however identify two 'features' of advanced thinking - acceptance of lack of closure, (ALC), and multiple interacting systems (MIS). Tasks which are complex and require an understanding of the interaction of two different systems involve MIS. Acceptance of lack of closure seems to be a general feature of the thinking of older children but not of younger children. It involves the ability to deal with intermediate stages in the solution of a problem which have indeterminate values. A good example of ALC is in the game 'Mastermind' in which the aim for the 'subject' is to guess which colour of pegs his opponent has put into four different slots. The subject is given feedback as to whether his guesses about colour and position are correct (black peg), whether colour only is correct (white peg) or whether colour is incorrect (no peg) but he is not informed about which feedback key corresponds to which of his response pegs: he may for instance be informed that two of his response pegs are the correct colours in the correct positions and that two are incorrect colours but he will not know which of his responses are which. According to Lunzer the younger subjects should find this kind of task difficult because the feedback they are given is indeterminate with respect to which of the four response pegs it relates to. Another example of ALC which will be discussed later in the thesis occurs with Denial ofthe Antecedent and Affirmation of the Consequent inferences in reasoning tasks. Young subjects cannot appreciate that the correct response to these argument forms is that no determinate conclusion can be drawn. Lunzer regards ALC as a 'feature' of more advanced thinking but not necessarily as a feature which depends upon a common underlying structure in the tasks for which it is required.

The main concern of Ennis in attempting to understand Piaget's theory was in trying to establish what predictions the theory made concerning the development of the ability to make deductive inferences from linguistic premises. From Piaget's description of formal thinking in terms of the operations of propositional logic and Inhelder and Piaget's assertion that "the principal novelty of the formal operational stage is the ability to reason in terms of verbally stated hypotheses and no longer merely in terms of concrete objects and their manipulation" (Inhelder and Piaget, 1958) it seems clear that Piaget intended his theory to apply to developmental changes in linguistic reasoning as well as more general changes in reasoning abilities. However, following a critical attempt to translate Piaget's theory about the logical structure of concrete and formal operational thought into empirically testable claims about verbal deductive logic ability at the different operational stages, Ennis (1975) dismissed Piaget's claims as either false, untestable or not about deductive logic.

One interpretation that can be made of Piaget's claim that children aged 7-8 to 11-12 years can handle class and relational logic but not propositional logic is that it entails that children in this age group can make valid deductions with linguistic problems expressed in terms of class and relational premises but not with arguments formulated in propositional logic (Osherson, 1975; Ennis, 1975, 1976; Brainerd, 1976-1977;). Thus subjects at concrete operations would be able to make correct logical deductions from linguistic expressions corresponding to class statements like "All A are B" but not until formal operations would they be able to make correct logical deductions from linguistic premises corresponding to propositional arguments such as "If p then q.". Empirical evidence however does not support this proposed difference in performance in reasoning with class and propositional arguments, showing unsystematic effects of the linguistic form of the premises on response (Hill, 1961; O'Brien and Shapiro, 1968; Roberge, 1970; Roberge and Paulus, 1971; Osherson, 1975). Under this interpretation Piaget's claims would be false. However Inhelder and Piaget's rejection of the linguistic criterion for determining a subject's operational level would indicate that this is probably an oversimplistic interpretation of the predictions of Piaget's theory vis a vis the development of linguistic reasoning ability.

A more theoretical problem which arises in trying to distinguish between the reasoning abilities typical of subjects at the concrete and formal operational levels is described by Brainerd (1976-77). He argues that Inhelder and Piaget's use of the logic of classes and relations as a model of concrete operational thought and propositional logic as a model of adolescent thinking is logically unsound since propositional logic is the most fundamental branch of logic (Quine, 1951) and provides the basis for both the logic of classes and the logic of relations. Brainerd argues that if the logic of classes and relations is a viable model of concrete operational thinking it is logically impossible (and presumably untenable as a developmental theory) that a more basic form of logic should be a model of a more mature developmental stage. In fact, according to Brainerd, logical considerations would dictate that the proposed order of acquisition of logical abilities would be the reverse of that proposed by Piaget since a subject who has mastered propositional logic should have the competence to master the logic of classes and relations.

It has already been mentioned that the logic Piaget uses as a model of formal operational thinking is not standard propositional logic but is more like propositional functional logic. Interpreting Piaget's propositional logic as propositional functional logic avoids Brainerd's criticism of the priority of Piaget's logical models but introduces further problems, since contemporary systems use propositional functional logic to interpret class inclusion statements like "All A are B". Propositional functional logic, or predicate logic, was developed to extend the logical analysis of sentences to a deeper level by attempting to deal with the problems of quantification but, given that the concrete operational subject is supposed to be able to deal with quantified statements like "All A are B", it becomes increasingly difficult to understand exactly what distinction Piaget intended between the logic of concrete and formal operations. Ennis (1975, 1976) found Piaget's use of a class inclusion example to demonstrate propositional implication in "Traite de Logique" (1949) particularly confusing for anybody trying to establish a useful distinction between concrete and formal logic and operations. It is this difficulty in distinguishing the logic of concrete and formal operations that led Ennis to argue that, under this interpretation, Piaget's theory is untestable.

The reasoning that Inhelder and Piaget describe in their hypothesis testing experiments involves reasoning from observed instances to general rules. This reasoning, as Ennis (1975) points out, is more like inductive reasoning than deductive reasoning. This was what Ennis meant when he objected that, according to one interpretation, Piaget's claims are not about deductive logic. Ennis claims that Piaget's logic goes from observed cases to general rules, which is an invalid deductive procedure. Piaget was attempting to specify how subjects come to abstract, from repeated exposures to instantiations of particular logical relationships in empirical situations, the specific combinations of elements corresponding to particular logical relationships. It seems plausible to suggest that although logically invalid, this procedure might be regarded as valid psychologically since the only way that a subject can come to propose that rules are general rather than domain specific is by extrapolation beyond observed instances.

Because of the difficulties Ennis found in establishing how Piaget's developmental theory was related to empirical findings about the acquisition of verbal deductive logical competence Ennis proposed an alternative framework for studying competence in reasoning with verbal premises. Ennis's framework takes cognizance of three different factors which have been found to influence acquisition of logical competence - logical principles, complexity and content. Ennis's system allowed for the acquisition of different logical principles at different ages in line with the empirical finding that valid principles e.g. detachment (Modus Ponens) and contraposition (Modus Tollens) are solved at an earlier age than invalid principles e.g. conversion (Affirmaton of the consequent) and inversion (Denial of the antecedent). This contrasts with Piaget's theory which entails synchronous development of all principles relevant to the particular logical system used to characterise the developmental stages.

Ennis maintained the distinction between logical principles of class, propositional and propositional functional logic in his framework but did not equate these different logics with particular developmental stages as Piaget did. He acknowledged that empirical studies might reveal no developmental differences at all between performance on logical principles from these systems. Ennis quotes the results of Flener's (1974) study which found no difference in the reasoning abilities of adult subjects with principles from propositional and propositional-functional logic.

Ennis's framework for studying logical competence acknowledges the substantial influence of content on conclusions drawn in reasoning tasks. His content dimension includes five elements: premise disbelief, commitment to a conclusion, symbolization, unfamiliarity and abstractness.

Ennis's framework also has a complexity dimension. More complex problems will be solved at a later stage than less complex problems.

The three different factors are not independent: the relative difficulty of different logical principles, for example, can be seen as partly attributable to their relative complexity, e.g D.A. is more difficult than M.P. because it includes negation, one of the factors which contributes to the complexity of a problem. Other factors contributing to the complexity of a problem include the number of connections, intricacy of argument, unrelatedness of content, nonstandard order and irrelevant material.

Despite Ennis's rejection of Piaget's theory as irrelevant in explaining the acquisition of deductive reasoning with verbal premises Piaget himself undoubtedly considered that his theory was relevant to reasoning with self-contained verbal premises: "propositional operations....are necessary for formal reasoning on hypotheses stated verbally" (Piaget and Inhelder, 1969 p.135). "The advent of formal operations brought about a major change in the subject's understanding of language and hence his ability to reason verbally" (Piaget and Inhelder, 1966). This was regarded as only one of the intellectual advances of the formal subject but the intellectual advances were of course presumed to be related by their common underlying stucture - the integrated structure of the combinatorial system with its transformational (INRC group) properties. Concrete proof that Piaget considered his theory relevant to explaining performance on verbal reasoning problems (if further proof is required) comes from his reply to a direct question about verbal deductive ability put to him by Suppes at a conference and quoted by Ennis (1975). Suppes asked him to explain his experimental observation that 6,7 and 8 year olds were very good at making simple Modus Ponens and Modus Tollens inferences. Presumably Suppes, like Ennis, felt that this facility with propositional logic inferences was a direct contradiction of Piaget's claims. The very fact that Piaget did not dismiss the question, arguing that his theory was not intended to cover purely linguistic reasoning, indicates that it was intended to do so.

Piaget explained in his reply how the distinction between the reasoning of concrete and formal operational subjects would be evident in purely linguistic problems:

"The problem is to know how the child reasons. Does he reason by means of situations that he can evoke or imagine, or does he reason by means of combinations of terms." (Piaget, 1967, p.277)

Although Ennis states that Piaget does not specify which subjects, concrete or formal, reason in which way, it is obvious that Piaget means that the concrete operational subjects reason by 'evocation of situations' and the formal operational subjects reason by 'combinations of terms'. There is no reason for Piaget to regard Suppes' evidence as running counter to his theory since he acknowledges that:

"it is possible to get correct reasoning about simple propositions as early as the 7-8 year level provided these correspond to sufficiently concrete representations." (Inhelder and Piaget, 1958, p. 252)

Suppes and Piaget seem to be in agreement that 7-8 year olds can give correct responses on natural language inference tasks provided that the content of the premises is sufficiently concrete (Piaget) or familiar to the subject (Suppes). Piaget does not regard the correct responses of the concrete subjects as evidence of precocious formal reasoning i.e. reasoning based upon an analysis of the formal stucture of the problem. Piaget always claimed that a child's understanding of language is constrained by the nature of the cognitive operations available to him. The concrete operational subject's understanding of language is constrained by the concrete or 'reality-based' nature of the operations available to him: with sufficiently concrete problems these operations can generate correct responses. The formal operational subject's response though is presumed to be generated by a qualitatively different analysis of the problem based upon the formal structure of the problem.

Although Piaget acknowledges that reasoning with simple propositions is possible at concrete operations "provided these correspond to sufficiently concrete representations" (Inhelder and Piaget, 1958) he does not discuss exactly what a "sufficiently concrete representation" might be. This leaves him open to criticisms of untestability since correct propositional reasoning by concrete operational subjects could be attributed to reasoning with sufficiently concrete representations while failure to respond correctly would indicate that the problems were not "sufficiently concrete". The distinction Piaget makes between the different means of reasoning of the concrete and formal subject is apparently similar to the distinction made in modern information processing theories between reasoning carried out by means of imaginal and propositional representations. The difficulties described in Part 1 in finding clear empirical evidence for the different representational modes are also relevant to Piaget's distinction between the verbal deductive abilities of concrete and formal operational subjects in terms of their "means" of reasoning.

The theoretical and empirical problems encountered in trying to establish the relevance of Piaget's theory to explanations of deductive reasoning ability and difficulties in trying to formulate empirically testable predictions of the theory have led others as well as Ennis to abandon Piaget's theory as a model for investigating deductive logical competence and its acquisition (Wason, 1977; Braine, 1978; Johnson-Laird, 1982, 1983). Besides the problems discussed already Piaget (1972) introduces further difficulties in distinguishing concrete and formal reasoning abilities with his admission that formal operational subjects are more likely to demonstrate their formal reasoning in problem areas with which they are familiar or in their areas of specialism. This modification was made in the light of evidence from many studies of the failure of supposedly formal operational subjects to respond correctly on a variety of formal reasoning tasks and significant improvements in performance when the content of the problem was concrete rather than abstract (Kodroff and Roberge, 1975; Evans, 1972; Wason, 1966, 1968, 1977; Wason and Johnson-Laird, 1972). However it introduces a further lack of testability into the theory since it would not be clear whether a failure to find formal operational thinking should be attributed to the lack of formal ability or to testing the subject in an unfamiliar content area. Changing the content of the problems until a suitable content area is found does not seem to be a very convincing method of demonstrating formal reasoning!

As with other putative formal operational abilities there is empirical evidence of substantial improvements in performance during early adolescence on tasks involving propositional logic (Shapiro and O'Brien, 1970; Roberge and Paulus, 1971; Taplin et al, 1974; Staudenmayer and Bourne, 1977; Kuhn, 1977; Roberge and Flexer, 1979, 1980). The results of studies of interrelationships between performance on propositional logic tasks and other formal operational tasks however have been more difficult to interpret (Kuhn, 1977; Kuhn and Angelev, 1976; Roberge, 1976). Roberge (1976) found evidence of significant improvements in performance on both a propositional logic task and a combinations problem at adolescence but found no correlation between scores on the two tasks. Roberge and Flexer (1979) found no significant correlation between scores on a propositional logic task and a combinations task for 8th. graders (mean age 13.5 years) or adults (22-48 years) but did find significant correlations between scores on the propositional logic test and a proportionality test, another formal operational task, for the adolsecents and adults. Roberge and Flexer (1980) found a significant correlation between performance on a control of variables task, another standard formal operational task, and a propositional logic task for grade 6 subjects (11.9 years) but not for grade 8 subjects (14.1 years).

The propositional logic tasks used as formal operational tasks typically require the subject to deduce conclusions from propositional premises: Roberge and Flexer, for instance included premises such as "If there is an A then there is a B. There is an A" and "Either there is an A or there is a B.". However other kinds of propositional logic tasks have also been regarded as requiring formal operational ability. Johnson-Laird and Wason (1977) and Wason (1977) for instance have argued that subjects at the stage of formal operations should have the competence to solve their selection task since this task involves a knowledge of propositional logic and hypothesis testing. Kuhn (1977) proposed that a task involving the conditional interpretation of empirical evidence was essentially a formal operational task.

Not only is the evidence concerning interrelationships between propositional logic tasks and the standard formal operational tasks like the isolation of variables and combinations tasks inconsistent, so too is the evidence concerning interrelationships between different propositional logic tasks. Roberge and Flexer (1980) reported no significant correlation between scores of 8th. grade subjects (14 years 1 month) on two propositional logic tasks, the propositional logic argument test, which required deductions like that described above, and the propositional logic statement test which involved verification of statements expressing a propositional logic relationship but the correlation at grade 6 (11 years 9 months) between scores on the two tasks was significant. Roberge and Flexer argue that this supports Kuhn's (1977) view that syllogistic reasoning is not dependent upon the emergence of formal operations. Interestingly Kuhn herself found a significant correlation between scores on a "conditional reasoning" task (a conditional verification task) and a propositional reasoning task involving categorical and conditional syllogisms at grade 8 (mean age 13 years 10 months) but no significant correlation at grade 6 (mean age 11 years 11 months). Inconsistent results such as these make it difficult to draw any firm conclusions about relationships between different propositional logic tasks and about whether particular tasks are formal operational. The significance of Kuhn's correlation between the two propositional logic tasks for grade 8 subjects, for instance, does not seem to support her conclusion that syllogistic reasoning does not require formal operations since the rationale for correlational analyses is that scores on tasks requiring formal operational thinking should be significantly related only for subjects at that stage since the set of abilities required to solve formal operational tasks is mediated by an organized, integrated structure. If Kuhn's claim is correct propositional inference tasks of the type which have been used as tests of formal operations are clearly inappropriate. The "substantial differences" Roberge and Flexer found in correct response on the two propositional logic tasks led Roberge and Flexer to conclude that:

"any assessment of adolescents' comprehension of propositional logic is highly dependent upon the precise nature of the problems used to measure this ability and that it is meaningless to speak of propositional reasoning ability per se."

Inconsistency of response by the same subjects on different propositional reasoning tasks (Evans 1978, 1980) and even by the same subjects on the same task (Taplin and Staudenmayer, 1973) was also noted in Part 1 where it was adduced as evidence against the rationalist approach to reasoning. Since Piaget's theory is a rationalist theory the inconsistent results on studies of performance on, and interrelationships between, propositional and other formal reasoning tasks has also been adduced as evidence against Piaget's theory.

However the more recent interpretation of the rationalist position in which a distinction is made between logical competence and performance has also been applied to Piaget's theory by Flavell and Wohlwill (1969). According to this view Piaget's description of operational structures constitutes a model of logical competence (Brown and Desforges, 1978; Boden, 1979; Falmagne, 1980; O'Brien and Overton, 1982). It has been argued that if Piaget's theory is understood as a model of logical competence it is incomplete as a theory of reasoning since it does not incorporate an account of the processes involved in utilizing the logical knowledge characterized in the competence component (Brown and Desforges, 1977). Although from the point of view of their logical structure certain tasks might be within the competence of the formal operational subject, a complete account of the responses subjects give in reasoning tasks also requires an account of the deviation of responses from the predictions of the competence model. Performance factors can be identified which account for the poor performance of adolescents and adults on propositional reasoning tasks, inconsistent response on different tasks which are nevertheless related in terms of their underlying logical structure, and inconsistent results on studies of interrelationships between such formal operational tasks. These include the information processing capacities of the different subjects, differences in the complexity, structure and content of the tasks, and different methodological procedures. Piaget's concentration on the logical structure of formal thought, his failure to give an adequate account of how the formal subject utilises this logical structure in solving particular problems and his failure to recognise that different aspects of tasks other than their logical structure can determine how well subjects perform on tasks make Piaget's theory as an account of the acquisition of reasoning abilities if not incorrect at least incomplete.

Falmagne's framework for investigating logical competence

Falmagne (1975, 1980) offers a very useful perspective on the development of logical competence. While Piaget regards propositional logic ability (in the sense of reasoning with verbal premises) as just one of the abilities which acquisition of the combinatorial structure of formal operations gives rise to, Falmagne distinguishes propositional reasoning studies and Piagetian studies as two different traditions in the approach to the development of logical competence. Falmagne points out that the two different traditions have focused upon different aspects of the child's logical development and claims that the extent to which these aspects are related is an empirical question and one which has not yet been rigorously addressed. The different traditions have dealt with different tasks and consequently implicate different subsets of cognitive skills. Within the propositional tradition the child is regarded as a logician attempting to make valid inferences by relying solely upon the formal properties of the premises and disregarding their content. Within the Piagetian tradition the child is regarded as a scientist attempting to discover and test general laws about the world. Falmagne herself was concerned with developing an account of the acquisition of logical competence with linguistic problems, an area she feels has been neglected due to the impact of Piaget's contention that pre-adolescents are unable to make propositional deductions.

Falmagne regards tasks within the propositional tradition as more strictly logical than tasks within the Piagetian tradition since the propositional tasks focus on the child's facility with the deductive procedures of the logician while the Piagetian tasks involve logical deductive reasoning as only a component part. The typical Piagetian task, unlike propositional reasoning tasks, also involves the subject's ability to generate and test hypotheses, an ability which involves inductive rather than deductive reasoning. The subject is required to observe empirical events, encode them in an appropriate way and abstract from this encoding a general rule or hypothesis which he can test.

Although the information encoded from linguistic and non-linguistic problems may be represented in similar ways, the way in which the information is extracted from the different types of problem must be different. (It will be remembered from the discussion of propositional representation that both linguistic and non-linguistic information can be represented propositionally.) Falmagne regards propositional reasoning as dependent upon the same kinds of representations and operations proposed in more general models of language comprehension and use, whereas the scientific reasoning tasks require processes involved in the encoding of empirical information. Falmagne proposes that both propositional and scientific tasks can be dealt with either formally, in which case the logical form of the problem would be extracted, or by a variety of content-based procedures. Although the propositional reasoning tasks are more strictly logical than the empirical tasks within the scientific tradition it is not clear whether a formal treatment of linguistically based problems would be more or less likely than a formal treatment of empirical problems. On the one hand the logical form of a verbal problem is more perspicuous than the logical form of an empirical problem. One reason for this seems to be that empirical states of affairs do not instantiate unique logical relationships. On the other hand the ambiguities associated with the formalization of natural language expressions would be absent in formalizing information in non-linguistic situations. It might be expected that empirical problems are more amenable to solution in terms of concrete modes of reasoning and verbal problems in terms of formal representations since these mappings between problem and representation are more intuitive than other possible mappings.

The alternative representations that Falmagne envisages to the formal approach to reasoning include reasoning with spatial/imaginal representations (Trabasso et al, 1975; Huttenlocher, 1968), protypical/analogical representations (Johnson-Laird, 1975; Johnson-Laird and Steedman, 1978; Guyote and Sternberg, 1978), Venn/Euler diagram representations (Revlis, 1975; Erickson, 1974).

Falmagne distinguishes between two different types of logical competence which are defined in terms of the different modes of reasoning. A subject is logically competent in the "strong" sense to the extent that he can generate correct inferences from a formal representation of the problem. Logical competence in the weak (or operational) sense requires only that subjects can generate correct inferences on a variety of inference tasks regardless of how these inferences are generated. Correct inferences could be generated by means of any of the alternative representational formats. Falmagne proposes that logical competence in both senses develops througout childhood and adulthood. Falmagne's distinction between weak and strong logical competence is similar to Piaget's distinction between concrete and formal operational thinking in that correct response by concrete operational subjects on linguistic reasoning problems is presumed to be generated not by an analysis of the logical structure of the problem but by means of concrete representations or 'evoked images'. Although Falmagne did not identify the different types of logical competence in terms of different developmental stages as Piaget did, she did propose that an important aspect of logical development was the increasing availability of different representational formats for solving problems and particularly the increasing availability of the formal mode for an increasingly wide range of problems with increasing age. For Piaget the essential difference between the reasoning abilities of concrete and formal operational subjects is the difference between the structures and operations available at the different stages. Falmagne draws the distinction between the reasoning abilities of younger and older subjects in terms of the availability of different representational formats for carrying out the deductions at different ages.

Falmagne's theory is more flexible than Piaget's in allowing that younger subjects, although more likely to use concrete based modes of representation to solve problems, can and do make purely formal deductions. She also argued that although the formal mode is increasingly available for the older subjects they will not necessarily solve problems formally even when the formal mode is available because it may be more expedient for a variety of reasons to use a concrete mode of representation.

Falmagne's definition of logical competence is different from other definitions (Osherson, 1974; Braine, 1977) in that she includes not just knowledge of logical relations and inference rules but also the ability of the subject to translate the natural language input in terms of an appropriate representational format to which inference rules can be applied. Other researchers would attribute the encoding of the premises to the linguistic encoding component of the model of reasoning, part of the performance component, rather than to the competence component. The encoding of the problem into a suitable representational format, or functional representation as Falmagne calls it, can be regarded as a mapping between the problem and the relevant available structures and representations in long term memory. The nature of this mapping is assumed to depend on the logical form of the problem, the nature and content of material and the availability of cues as to the logical form of the problem. A formal representation of the problem is a mapping between the problem and logical relationships in long term memory. When the functional representation is not formal the structural properties of the representation are largely determined by the content of LTM. Falmagne's account of the formalization of natural language input into a functional representation highlights her view of logical reasoning as closely integrated with the general processes carried out in language comprehension.

It will be remembered from the discussion earlier on that Falmagne does not regard logic as absent from reasoning carried out by means of the alternative (non-formal) representational modes but rather she claims that logic is hidden in the formalization process. By this she seems to mean that the choice and use of one particular representational format rather than another to represent the problem is determined by the subject's stored knowledge of the logical properties of that representation and his ability to identify a particular representational format as relevant to the solution of the problem. The subject has stored in long term memory knowledge of the logical properties of the particular representational format and can use cues from the problem to recognise the problem as an example of that representational format.

If the concrete modes of reasoning incorporate logical knowledge it seems pertinent to ask why concrete representations are necessary at all and why the information is not just represented formally. Logical knowledge incorporated in concrete models of reasoning is not logical in the same way that logical knowledge incorporated in a formal representation is logical. Concrete representations represent the logical properties of statements in an explicit "graphic" way. Whether this explicit representation of logical properties should be regarded as logical knowledge is an equivocal issue and is obviously related to the longstanding philosophical problem of what is to be counted as logical knowledge. Falmagne regards logical knowledge as knowledge about linguistic structure and knowledge about the real world "to the extent that language itself is semantically grounded in the empirical world to which parts of it refer" (Falmagne, 1980, p.176). Falmagne believes that in choosing a specific concrete representation of a particular sentence, the subject is demonstrating his understanding of the isomorphism between the logical properties of that sentence and the logical properties of the concrete representation; knowledge of the logical properties of the concrete representation is stored in long term memory. This reflects her belief that logical knowledge derives from an awareness of how language maps on to the real world. This contrasts with Piaget's view since for Piaget language is not central to the acquisition of logical knowledge. For Piaget logical operations are constructed by the internalization of actions upon objects in the environment and the integration of these into a coherent interconnected system; logic is both constrained by and is the result of our cognitive processes. Johnson-Laird's proposal that logic is an emergent property of manipulations of 'models' -concrete representations - seems to be dependent upon his view that language is interpreted with respect to a determinate model. The logical properties of a sentence can only be derived from an exhaustive consideration of all the distinct models which accurately represent the truth of the sentence. For Falmagne the concrete representation of a problem which the subject chooses is a canonical representation incorporating its logical properties i.e. an abstraction from all models compatible with the truth of the sentence. Falmagne would argue that the choice and use of a particular model as an appropriate representation of a sentence reflects the subject's knowledge of the structural properties of the model and its relevance in encoding the sentence. It is interesting in this respect that Falmagne does not account for how the child would come to have this knowledge of the logical properties of concrete representations although it seems likely that she would invoke a Piagetian explanation since she regards many of the basic assumptions of her framework as "in a general sense, entirely congruent with the constructive epistemology represented in Piaget's theory" (Falmagne, 1980, p.181).

Falmagne argues that her approach to the study of the acquisition of verbal deductive logical competence is "not seen as incompatible with Piaget's theory". Like other recent theorists within the information processing tradition, she regards Piaget's theory of formal operations as a structural description of the adolescent's logical competence and, as such, she regards Piaget's theory as incomplete since he fails to account for discrepancies between the predictions of the logical model and the observed behaviour of subjects in reasoning tasks, other than to invoke the concept of horizontal decalage to explain why facility with concepts which come from different content areas but which have the same underlying structure does not emerge concurrently.

Central to Falmagne's framework for studying the acquisition of logical reasoning abilities is her view of the ability to make logical deductive inferences as closely integrated with the structures and representations proposed in theories of language comprehension and in general theories of cognitive processes within cognitive psychology. Her view that deductive reasoning with linguistic premises can be explained in terms of the same structures and processes as those proposed in theories of sentence comprehension highlights some of the inadequacies of Piaget's theory since Piaget did not consider specifically linguistic aspects of logical abilities. It makes her approach to verbal deduction more useful than Piaget's approach although "not incompatible" with it. Falmagne's approach offers a way of giving more substance to Piaget's claims about the different "means" of reasoning of younger and older subjects. Falmagne distinguishes between a general account of the development of logical competence with respect to the development of other aspects of cognition and an account of more specific "microdevelopmental changes" which may determine facility with particular patterns of inference or logical relationships. Piaget has provided a general theory of the changes in cognitive structure which take place around adolescence but has not been concerned with exploring the details of this change.

It is the objective of this thesis to explore the details of how the ability to understand and reason with universally quantified statements and general conditional statements is acquired. From Piaget's account of the development of logical thinking it is difficult to make clear, unambiguous predictions concerning the development of linguistic reasoning abilities concerning these statements. However because of the centrality of the conditional to accounts of reasoning it is important to establish the extent to which children are capable of understanding this relation.

Research on language comprehension and inference has been concerned with providing models of performance on different comprehension and inference tasks in terms of the different cognitive processes implicated in solving these tasks. Performance is typically assessed in terms of error rates or latencies. Within the Piagetian tradition any deviation from correct response on a formal inference task has been considered as evidence against Piaget's theory. A more fruitful approach to providing a theory of reasoning ability is to attempt to characterise the demands of the different tasks and, by presenting these tasks to subjects of different ages, to establish the age at which facility on the different tasks is acquired. This would allow the developmental course of the acquisition of the component abilities implicated in the solution of these tasks to be elucidated. It is tempting to suppose that certain tasks assess comprehension of sentences while others assess inference from these sentences. In practice it is impossible to assess comprehension and inference independently since the interpretation of a sentence determines to a certain extent the permissible inferences while adequate comprehension of a sentence requires that certain inferences be made. The difficulty in characterising comprehension and inference separately was discussed in part 1. Rather than becoming embroiled in the debate about the extent to which different tasks involve comprehension and inference a more fruitful approach to providing a theory of reasoning ability is to attempt to characterise the demands of the different tasks and, by presenting these tasks to subjects of different ages, to establish the age at which facility on the different tasks is acquired. This would allow the developmental course of the acquisition of the component abilities implicated in the solution of these tasks to be elucidated. The different tasks used in this thesis include a verification task, an evaluation task and syllogistic reasoning tasks.

Although verification tasks are often referred to as simple reasoning tasks they do not involve deductive reasoning but require the subject to say whether a sentence is a true or false description of a particular empirical situation. However as Evans (1982) states "these tasks provide information about people's ability to represent and process linguistic structures that are involved in bona fide deductive reasoning problems".

As explained in part 1 the rationalist proposal that inference tasks reflect the interpretation of linguistic premises on which inference is then carried out in accordance with logical principles led to an abundance of research using inference tasks in an attempt to establish how sentences are understood. The evaluation task is a kind of inference task which has been used to assess comprehension of the linguistic premises and particularly the conditional. The subject is presented with a linguistic statement and either has to say whether particular exemplars are compatible with that statement or whether particular exemplars make the statement true. This task involves purely linguistic premises and consequently from a Piagetian viewpoint should be more difficult than the verification task. The syllogistic reasoning task has also been used to assess how statements are understood. The subject is presented with a major and a minor premise and is required either to make an inference or to assess the validity of an inference. The subject's response is considered to reflect his interpretation of the major premise.

Because of Piaget's claim about the logical abilities of subjects at the concrete and formal operational stages of development performance on both class and conditional versions of the tasks was compared. Experiments 1, 2, 3, and 4 were concerned with the verification of statements of this type; experiment 5 was an evaluation experiment while experiments 6 and 7 were studies of syllogistic reasoning ablities.

One aspect of the development of comprehension and inferential abilities which Piaget did identify but which information processing theorists have, until recently, largely ignored is the fact that the development of logical abilities is, in the first place, tied to concrete objects and relations. The subject's ability to make inferences initially depends upon the presence of concrete objects. It seems feasible to propose that comprehension and inference tasks which are related to concrete objects and relations will be solved at an earlier age than those which require reasoning with self-contained linguistic premises. Further since class statements like "All A are B" map in a more intuitively obvious way than propositional statements "If p then q" onto empirical objects any distinction between the comprehension of class and conditional statements would be more likely to occur in comprehension tasks involving interpretation of empirical information rather than in purely linguistic tasks.

CHAPTER 2

UNDERSTANDING UNIVERSALLY QUANTIFIED STATEMENTS

In "The Early Growth of Logic in the Child" (1964) Inhelder and Piaget describe their extensive investigations of the developmental mechanisms of the classification and seriation operations which characterise the concrete operational stage of the child's cognitive development.

Of particular interest to this thesis was an experiment which they conducted on the young child's understanding of class inclusion and comprehension of the quantifiers "all" and "some". In this experiment the child was presented with a collection of counters ranging in number from 8 to 21 and consisting of blue squares, blue circles, and red squares (see Figure 2.1). He was then asked the following questions about the collection:

(1) Are all the circles blue?
(2) Are all the red ones square?
(3) Are all the blue ones circles?
(4) Are all the square ones red?

Although an adult would immediately respond "yes" to (1) and (2), and "no" to (3) and (4), Inhelder and Piaget found that 5 year-olds found the task quite difficult and responded correctly to only 67.5% of such questions. 7 year-olds were correct on 85% of questions, while by 9 years old 94% of responses were correct.

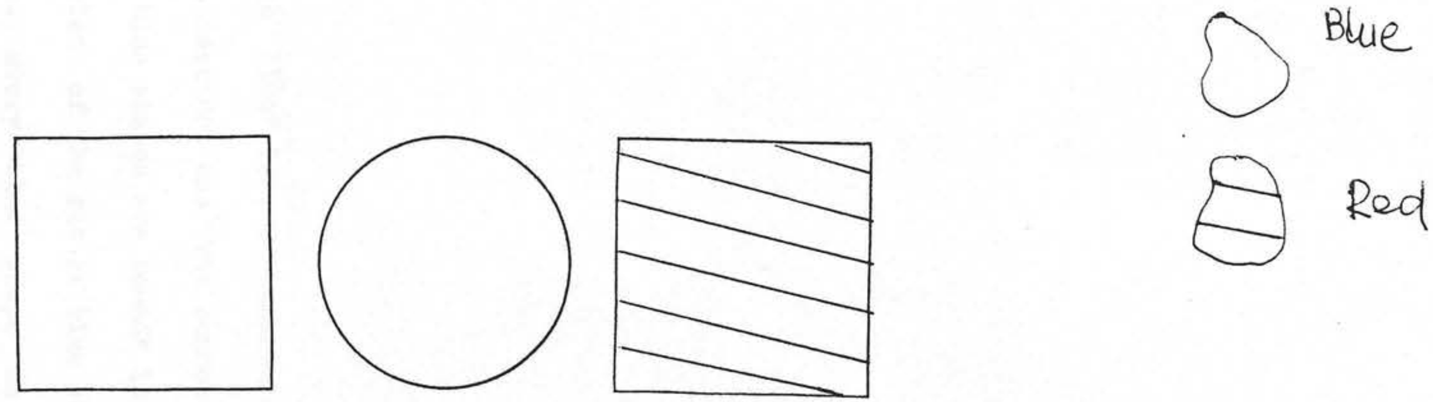
Inhelder and Piaget argue that the children's errors on this kind of problem are due to "false quantification of the predicate". Rather than interpreting the universal affirmative proposition:

(5) All the circles are blue.

as the inclusion relation:

Figure 2.1

Picture of Shapes in Class Inclusion Task



(6) "All the circles are some of the blue things."

the young children misinterpret (5) as:

(7) "All the circles are all of the blue things."

With the collection of shapes in Figure 11 this interpretation would clearly lead the subject to give the wrong answer i.e. that all the circles are not blue (i.e. are not all the blue things) because there are some blue squares.

Inhelder and Piaget argue that the difficulty that the young child has in quantifying the predicate arises from more fundamental cognitive difficulties in operational classification. A child who has mastered operational classification understands that the statement: "All the roses are flowers" is true because the extension of the set of roses defined by the quantifier "all", namely every member of that set, is marked by the intensive characteristic of being a flower. In a similar way the concrete operational child would understand that: "All the circles are blue" is true because every member of the set of circles in the collection has the common characteristic "blue". Similarly, "All the blue shapes are round" is false because it is not true that the extension of the set of blue shapes designated by the quantifier "all", i.e. every blue shape in the collection, is marked by the common characteristic "round" since some blue shapes are not round.

The adequacy of the false quantification of the predicate hypothesis (the F.Q.P. hypothesis) has been questioned both on theoretical grounds and empirical grounds. Kneale and Kneale (1962) point out that quantifying predicates raises problems of logical formulation. Bucci (1978) has criticized the F.Q.P. hypothesis on empirical grounds. She points out that young children incorrectly reject sentence (1) as false not just "because there are blue squares" but also "because there are red squares." Bucci argues that the F.Q.P. hypothesis cannot explain the "red squares" type of error since the presence of red squares would be irrelevant in a comparison of "all the circles" with "all the blues".

Although Bucci claims that the F.Q.P. hypothesis cannot account for the "red squares" type of error Inhelder and Piaget did offer an explanation of such errors in young children. Inhelder and Piaget claimed that the preoperational child cannot yet reason in terms of classes because he cannot differentiate between the intension and extension of a class. He reasons instead in terms of 'non-graphic collections', which are prelogical collections of objects with similar characteristics. Rather than regarding "all" as a quantifier defining the extension of a class, the child at the pre-operational stage regards "all" as referring to the properties of a collection of objects. In the words of Inhelder and Piaget:

"the collection as a whole is exclusively and entirely blue and made up of circles just as a particular object must be entirely blue as well as circular to be called a 'blue circle.'"

In assessing the truth of:

"All the circles are blue."

the subject compares the two collections of round elements and blue elements to decide whether these two sets coincide. Any element which is not blue and round, including red squares, would be a counterexample to this interpretation of the quantified statement.

Another hypothesis proposed to explain how quantified statements are interpreted is Chapman and Chapman's (1959) illicit conversion hypothesis. According to this hypothesis many subjects interpret statements of the type "All As are Bs" to mean that the converse "All Bs are As" is also true. Effectively the F.Q.P. hypothesis and the conversion hypothesis both assume that the subject misinterprets an inclusion relationship (A<B) as an equivalence relationship (A = B). The difference is that according to the F.Q.P. hypothesis this misinterpretation is not strictly logical since it is based upon the subject's manipulations of collections rather than logical classes. A subject who converts "All As are Bs" to "All Bs are As" apparently understands that the conjunction of the inclusion relation and its converse "All As are Bs and All Bs are As" is a logical equivalence relationship (A = B).

Revlin and Leirer (1980) have recently proposed an alternative form of the conversion hypothesis. They propose, rather counterintuitively, that in the process of comprehension the subject converts the sentence "All A are B" to "All B are A" and that this converted relationship is most salient in further processing during comprehension or reasoning tasks and hence in determining the subject's response. This hypothesis would presumably not be relevant in explaining the errors of young children however since the conversion operation is dependent upon the correct assignment of the inclusion relationship and it is this inclusion relationship which the younger children have not mastered.

Bucci rejects the F.Q.P. hypothesis and the illicit conversion hypothesis of how the subject-predicate distinction can be nullified, and proposes instead a "structure neutral" hypothesis which holds that, in interpreting universal affirmative propositions, subjects of certain ages and in certain circumstances, do not assign the subject-predicate distinction properly. The universal affirmative proposition "All the circles are blue" is encoded as an unordered string of content words without any structure "All, circles, blue". Unless further interpretation is imposed by the content or context of the statement or by re-encoding the sentence with further attention to its structure, the most likely interpretation of this structure neutral encoding will be something like "Everything is a circle and blue" or "They're all blue circles". The structure neutral interpretation is similar to the Piagetian interpretation which proposed that "All A are B" is understood to mean that the collection as a whole is "all A and B". Bucci presents evidence in her paper that even adults apparently interpret universally quantified statements in this way sometimes.

A subject using a structure neutral (S.N.) interpretation of sentences (1)-(4) would respond "false" to all 4 sentences since in every case there are shapes present in the collection of Figure 2.1 which would contradict the S.N. interpretation. In sentence (1) for instance, his interpretation "Everything is round and blue" would be contradicted both by the presence of blue squares and red squares. Bucci's S.N. hypothesis would predict that if young children do use a S.N. interpretation they should respond "false" even to sentences which are true. Performance on the task described by Piaget should be characterised by superior performance on false items compared with true items.

Although Bucci claims that the implications of this differential performance on true versus false items were not discussed by Inhelder and Piaget, this would also be a prediction of Piaget's "collection-as-a-whole" interpretation since this interpretation is similar to the S. N. hypothesis and Inhelder and Piaget do say (in

"The Early Growth of Logic in the Child", p. 71) that children find it easier to say that "All Bs are A" is false than to say that "All As are B" is true. Indeed the difference is evident in Piaget's experimental data. Inhelder and Piaget's data show that overall more subjects respond correctly to all false items than all true items although they do not say whether this difference is significant. This differential performance on true and false items would also be predicted by the F.Q.P hypothesis and the Chapmans' conversion hypothesis. The false statement "All blues are round" would be classified false under the F.Q.P. interpretation and the conversion hypothesis because of the presence of blue squares. The true statement "All circles are blue" would also be classified as false under both the F.Q.P. and conversion interpretations, because of the presence of blue squares.

Bucci reports an experiment - "The Display Task" - which was essentially a replication of Inhelder and Piaget's study of the comprehension of the quantifier "all". Bucci used three different age-groups: 37 children from 6 years 6 months to 8 years 8 months; 38 children from 11 years 5 months to 12 years 5 months and 28 adults.

The subject was presented with a card on which there were five or six small plastic blocks varying in size, shape and colour, and a sentence e.g. on one card there were yellow squares, blue squares and blue rectangles, and a sentence:

(1) On this card all the yellow blocks are square.

or (2) On this card all the blue blocks are rectangles.

The major differences between Bucci's and Inhelder and Piaget's tasks were:

(a) The set size: Bucci used 5 or 6 objects whereas Inhelder and

Piaget used 8 to 21 objects.

(b) The number of items: Bucci used 3 true and 3 false items: Inhelder and Piaget used only 2 true and 2 false items.

(c) The type of task: Bucci's task was a verification task: Inhelder and Piaget's task was a question/answer task.

Comparable data were available from Inhelder and Piaget's study only for the 6-8 year olds. The results for this age-group were similar: 84% of responses by Bucci's 6-8 year-olds and 85% of responses by Inhelder and Piaget's 6-8 year-olds were correct. 54% of Bucci's 6-8 year-olds and 47% of Inhelder and Piaget's 6-8 year-olds were correct on all items.

Bucci's results show that her predictions about differential performance on "true" versus "false" items was upheld. Performance on false items was virtually unchanged between groups. Just over 80% of 6-8 year-olds, 11-12 year-olds and adults answered all false items correctly. Performance on true items however was very poor with only 39% of 6-8 year-olds and 59% of 11-12 year-olds answering all the true items correctly. There was little difference in performance on true versus false items by adults: 88% of adults answered all true items and 82% of adults answered all false items correctly. There was a significant difference between older children and adults in response to true items (chi-square = 7.86, df = 1, $p<0.01$). The difference between younger children and older children's response to true items just missed significance (chi-square = 3.07, df = 1, $p<0.10$).

Bucci attributes the difference in performance on true and false items to a structure neutral interpretation of the universal affirmative proposition by most of the younger children (6-8 years) and many of the older (11-12 year-olds) children. However, as we have seen, this difference is also compatible with Piaget's F.Q.P. hypothesis and the Chapmans' conversion hypothesis.

There were several flaws in Bucci's experiment which could account for her results:

(1) Bucci used the term "rectangle" in problems where squares and rectangles were presented in the same collection. It is quite likely that most 11-12 year-olds and adults if not 6-8 year-olds know that "All squares are rectangles". If this is considered to be relevant in Bucci's experiment, statements like "All the blue blocks are rectangles" are actually true of a collection of blue and yellow squares and blue rectangles like that used by Bucci. This example would lead to errors on false items rather than true items, and would not affect the S.N. hypothesis. Consider though the converse of this statement which was presumably used in Bucci's study:

"All the rectangles are blue."

This statement is true only if one considers squares and rectangles to be disjoint classes. If the subject considers that his knowledge that "All squares are rectangles" is relevant, the sentence is clearly false since there are yellow rectangles (i.e. yellow squares) present. It is possible that this type of "error" contributed to the poor performance on true statements by the 11-12 year-old subjects since it it usually impressed upon subjects of this age that "All squares are rectangles". Since young subjects find it easier to respond to statements with shape rather than colour in the subject position and since all statements of the type Bucci used with the term "rectangle" in the subject position will be false if the fact that "All squares are rectangles" is taken into account, it is plausible that more "errors" will be made in classifying sentences which Bucci classifies as true (i.e. those with "rectangle" in subject position) than in classifying those which Bucci classifies as false (i.e. those with "rectangle" in predicate position). Any errors attributable to ambiguity associated with the term "rectangle" could easily be avoided by excluding that term from the experiment.

(2) Young children presented with 5 shapes (2 yellow squares, 2 yellow circles and a blue square) and asked whether "All the circles are yellow" is true might respond that it is false, the reason being that only 2 circles are yellow. Inhelder and Piaget report this kind of justification by younger children who seem to require that "all" refers to a larger number of items than two. Errors of this type would also lead to an inflated number of errors on true items compared with false items. It is possible that some errors on true statements were due to this type of error since Bucci used small set sizes. The small set size was used to obviate the possibility of errors due to careless inspection of the shapes. A set size of 8 to 9 objects would be more likely to avoid both pitfalls.

(3) Piaget reports that children produce more errors on both true and false items when the quantified term concerns colour (sentences 2 and 3) than when the quantified term concerns shape (sentences 1 and 4). Since Bucci does not mention this factor we do not know whether she controlled for it or not. If there were more true items than false items with a quantified shape term then this would lead to an increase in errors on true items over false items.

(4) As previously mentioned, Bucci's results were reported in terms of

percentage of subjects correct on all the true versus false items. It is not obvious why she reports her results in this way rather than in terms of mean correct response to "true" versus "false" tems. This latter method of analysis is more useful for comparing the results of different experiments since the former method depends to some extent on the number of items included. Obviously the more items the less likelihood of a subject being correct on all items of a particular kind although there is no reason why there should be a differentiation in correct response to "true" versus "false" items for any particular number of items.

Experiment 1 was carried out to discover whether Bucci's results could be replicated in a study which obviated some of the flaws in Bucci's study. The following modifications were made:

(a) No rectangles were used.

(b) Set sizes of 8 objects were used.

(c) Quantified terms concerning colour and shape were balanced across the "true" and "false" items.

(d) Analysis was performed both in terms of percentage correct response to "true" versus "false" items by subjects at a given age, and also in terms of percentage of subjects at a given age responding correctly to all "true" versus "false" items.

EXPERIMENT 1

THE INTERPRETATION OF UNIVERSAL AFFIRMATIVE PROPOSITIONS

METHOD

Materials

Twelve different 8" by 5" index cards were prepared for presentation to the subjects. On each card there were eight shapes of three different kinds, varying along the dimensions colour and shape. Three different shapes (triangle, circle and square) and three different colours (blue, red and green) were used but the attributes on any one card were binary. The relationship between the attributes of the coloured shapes on each card could be characterised as one of inclusion. A card on which there were blue circles, blue squares and green squares, for example, represented the inclusion relationship: "All the circles are blue." or "All the green shapes are square.". Twelve out of the thirty-six possible such relationships between the attributes were chosen in such a way that each colour and type of shape appeared four times over the twelve cards.

In addition to the shapes, each card had printed on it a universal affirmative proposition like the following:

(8) On this card all the circles are blue.

Six of the statements were true of the relationship depicted on the card, and six were false. The false statements were always the converses of statements which were true of the relationship depicted on the card. If, for example, it was actually true that all the circles on the card were blue the false statement used would be:

(9) On this card all the blue shapes are circles.

The nature of the quantified term, colour or shape, was crossed with truth value of the statement so that there were 3 true and 3 false statements with the quantified term concerning colour (TC and FC) and 3 true and 3 false statements with the quantified term concerning shape (TS and FS). Each colour and shape appeared once in each of these four different proposition types.

Seven practice cards were also prepared. 3 of these served as an introduction to the experiment and required the children to name the shapes on each of the cards - triangles on one card, squares on the second, and circles on the third. Each shape was presented in red, blue and green, and the child was also required to name the colours. This was to ensure that the child was familiar with the shapes and colours used. The next two practice cards contained two different types of shape and a statement to be verified (one true and one false) and the last two practice cards contained three different shapes and a statement to be verified. The practice session familiarised the subject with the task and ascertained whether the subject could read all of the words used in the experiment.

SUBJECTS

Subjects were 20 Primary 2 children, and 20 Primary 7 children from an Edinburgh Primary School. Mean age of the Primary 2 children was 6 years 2 months (with a range of 5 years 5 months to 6 years 8 months), and for the Primary 7 children, 11 years 0 months (with a range of 10 years 5 months to 11 years 6 months). Subjects were chosen randomly by their form teachers as of average to above-average ability.

PROCEDURE

Each child was interviewed individually by the experimenter and the task took between ten and fifteen minutes. With the aid of the practice material it was explained to the subject that the task involved colours and shapes, and that he was going to have to say whether some sentences about colours and shapes were true or false.

The twelve experimental cards were then presented to the subject one after the other in random order. The child was told to look carefully at the coloured shapes on each card, to read the sentence and then to say whether the sentence was true or false. Some of the Primary 2 children found it easier or more natural to say "right" or "wrong" instead of "true" or "false" as they were more familiar with these terms. As it was felt that the two types of judgement did not differ, the children were allowed to use "right" and "wrong" if they wished. No child had any difficulty with the procedure.

RESULTS

Table 2.1 shows the percentage of correct responses to the different types of proposition for subjects at different grades. Overall, Primary 2 children were correct on 71% of responses and Primary 7 children were correct on 95% of responses. A t-test on independent means revealed a significant effect of grade ($t=6.87$, $df=38$, $p<0.001$). The Primary 7 children show essentially mature performance on this task, but Primary 2 children still make errors.

Contrary to Bucci's predictions, there was no difference between correct response to true items and false items for children at either grade. 71% of true responses and 70% of false responses by Primary 2 children, and 94% of true responses and 95% of false responses by Primary 7 children were correct. In contrast with Bucci's results

TABLE 2.1

PERCENT CORRECT RESPONSE TO DIFFERENT TYPES OF PROPOSITION FOR SUBJECTS AT DIFFERENT GRADES

	P2	P7
TC	58	92
TS	83	97
FC	60	95
FS	80	95
MEAN TC+TS	71	94
MEAN FC+FS	70	95
MEAN TC+FC	59	93
MEAN TS+FS	81	96
OVERALL MEAN	71	95

TC: True statement with colour subject
FC: False statement with colour subject
TS: True statement with shape subject
FS: False statement with shape subject

TABLE 2.2

NUMBER OF SUBJECTS AT EACH GRADE CORRECT ON ALL PROPOSITIONS OF A PARTICULAR TYPE

	P2	P7
TC	5	15
TS	13	18
FC	5	17
FS	12	17
TC+TS	3	13
FC+FS	5	14
TC+FC	1	12
TS+FS	9	15
MEAN	1	12

TC: True statement with colour subject
FC: False statement with colour subject
TS: True statement with shape subject
FS: False statement with shape subject

which showed a trend towards a significant difference between the responses of the younger and older children only on true items, a significant difference was found between the responses of younger children (6 year-olds) and those of older children (11 year-olds) on both true ($t=4.29$, $df=38$, $p<0.001$) and false ($t=4.44$, $df=38$, $p<0.001$) propositions.

Since Bucci's analysis was in terms of percentage of subjects correct on all the true versus all the false items, a similar analysis was performed here. Since more items were included in this experiment fewer subjects would be expected to be correct on all items, but according to Bucci's predictions, a true/false differentiation should occur. Table 2.2 shows the percentage of subjects at each grade correct on all propositions of a particular type. There were 12 propositions to be verified in this experiment and only one 6 year old subject (5%) and twelve 11 year olds (60%) were correct on all propositions (compared with 54% for Bucci's 6-8 year-olds and 71% for 11-12 year olds). As regards the number of subjects correct on all true versus false items, 3 (15%) Primary 2 and 13 (65%) Primary 7 subjects were correct on all "true" items, while 5 (25%) of Primary 2 and 14 (70%) of Primary 7 subjects were correct on all "false" items. The difference in correct response between Primary 2 and Primary 7 children, as with the percent correct response analysis, was significant on both the true items (Chi-Square = 10.416, $df=1$, $p<0.001$), and on false items (Chi-Square =8.12, $df=1$, $p<0.005$) but there was no difference in correct response to true versus false items at either grade.

From Table 2.1 it can be seen that Primary 2 children gave more correct responses to items with shape subjects (81%) than to items with colour subjects (59%). A t-test for correlated means shows that this difference is significant ($t=2.81$, $df=19$, $p<0.01$). There is no such difference for Primary 7 subjects who performed well on propositions of both types: 93% of responses to colour items and 96% of responses to shape items were correct.

Table 2.2 shows that the shape/colour subject difference is also reflected in the number of Primary 2 subjects correct on all colour items (only 5%) compared with all shape items (45%). No such difference was found for Primary 7 children.

DISCUSSION

The results of experiment 1 do not support Bucci's prediction concerning response differences to "true" and "false" propositions. There was no difference in response on "true" versus "false" items for the 6 year-olds or the 11 year-olds, whether this was in terms of mean correct response to "true" versus "false" items or in terms of number of subjects at a given age correct on all "true" versus "false" items.

There was a significant effect of grade with 11 year-olds performing better than 6 year-olds overall. The effect of grade was also significant for both "true" items and "false" items. This indicates that the absence of a true/false differentiation in response was not due to an absence of errors by 6 year-olds. Indeed 6 year-olds in experiment 1 made more errors (29%) than the young subjects in Bucci's study (16%) although the errors were on both "true" and "false" items. The higher error rate for 6 year-olds than that for Bucci's subjects was probably due to the wider age-range used by Bucci who used children aged from 6-8 years.

Bucci's results did not uniquely support her stucture neutral hypothesis since both Piaget's F.Q.P. hypothesis and the Chapmans' conversion hypothesis would also predict a higher error rate on true compared with false items. The absence of a higher error rate in experiment 1 is thus incompatible with the S.N. hypothesis, the F.Q.P. hypothesis, and the Chapman's conversion hypothesis. Although all three hypotheses can explain errors on true items none of these can explain errors on false items. Of the hypotheses mentioned in the introduction only Revlin and Leirer's conversion hypothesis could explain errors on false statements. This conversion hypothesis proposes that in encoding quantified sentences subjects carry out a conversion operation on the subject so that the converse is prepotent in future processing. Thus the true statement "All the round shapes are blue" of Figure 21 would be converted to a false statement "All the blue shapes are round" while the false statement "All the blues are round" would be converted to the true statement "All round shapes are blue". All other things being equal, errors on true and false propositions would be equally likely. Although Revlin and Leirer's conversion hypothesis was not thought to be applicable to young children's responses on verification of quantified statements this type of conversion hypothesis can account for the similar error rates on "true" versus "false" items.

Despite Bucci's contention that the true/false differentiation which she found was not reported by Inhelder and Piaget, Inhelder and Piaget (1964) did report superior performance with false statements. They explained the errors on true statements in terms of the F.Q.P. hypothesis and also emphasised that some correct responses on false statements were probably correct for the wrong reason i.e. some subjects who correctly argued that "All B are A" is false were basing their responses not on the fact that there are some Bs which are not-a, but on the fact that Bs and As are not co-extensive. Inhelder and Piaget argued that errors on false items arise because of conversion of "All B are A." to "All A are B.". Although Inhelder and Piaget argued that interpreting the inclusion relation as its converse is similar to substituting equivalence for inclusion this is not true: the interpretation is one of reciprocal inclusion rather than equivalence. As we have seen interpreting a false inclusion statement (All B are A) as an equivalence would not lead to errors in classifying false statements.

Inhelder and Piaget proposed that the difference in response on true and false propositions would be found only for younger subjects. Bucci actually found a significant increase in correct response on true propositions between 11-12 year olds and adult subjects indicating that older subjects are also liable to misinterpret the true propositions. This suggests that the absence of a true/false differentiation in experiment 1 was not caused by using subjects who were too old to demonstrate the effect.

The difference in response of 6 year olds to items with shape subject (81% correct) compared with a colour subject (59% correct) supported Piaget's view that children of this age are still heavily influenced by the perceptual characteristics of the picture or array since they base their response on collections of elements rather than on an understanding of one class included in another. It seems that 6 year olds find it easier to group together items of a particular shape rather than items of a particular colour. It seems likely that the young subjects adopt a strategy of grouping together shape items rather than colour items even when it is not appropriate to do so. When the subject of the sentence concerns shape, this strategy is effective whether the sentence is true or false. If the sentence is true e.g. "All the circles are blue." the subject will group together the circles and will readily affirm that they are all blue: 83% of P2 responses on true sentences with a shape subject were correct. If the sentence is false e.g. "All the square shapes are red." the subject will group the squares together and realize that a blue square is a counterexample to the truth of the sentence: 80% of P2 responses to false sentences with shape subjects were correct. When the subject of the sentence concerns colour the strategy of grouping by shape is likely to lead to error with both true and false sentences. The subject who uses a 'grouping by shape' strategy in responding to a true sentence such as "All the red shapes are square." will tend to group the squares together and regard a blue square incorrectly as a counterexample to the sentence: only 58% of P2 responses to true sentences with colour subjects were correct. Grouping by shape would also lead the subject to regard a false sentence such as "All the blue shapes are round." erroneously as true since the subject would group the circles together and would find that these are all blue: only 60% of false sentences with shape subjects were correct. Either the child focuses on the shape whether or not this is the subject or predicate of the sentence or the child converts the sentence where necessary so that the shape is the subject.

It is interesting to note that Just (1974) found that adults in a sentence picture verification task verified both true and false sentences with a colour subject e.g. "All the red figures are round." 146ms. faster than sentences with a shape subject e.g. "All the round figures are red." when they were required to respond as quickly as possible. The difference disappeared when subjects were given ample time to respond. Just argued that since colour is processed better than shape in peripheral vision under time pressure subjects fixate items of a particular colour rather than items of a particular shape.

Donaldson (1978) reports an experiment in which what might be called 'conceptual salience' rather than perceptual salience determined response on a verification task. The problem concerned four garages and a set of either three cars, in which case the three cars were in garages with one garage empty, or five cars in which case the four garages had cars in them with one car left over. Children under seven years were required to judge whether the sentences:

(1) "All the cars are in the garages."

(2) "All the garages have cars in them."

were true or false in the three car and five car cases. In the three car case, (1) was true and (2) false while in the five car case (1) was false and (2) true. Subjects seem to focus on 'cars in garages' saying that both (1) and (2) are true with five cars but that they were false with three cars. It could be argued that their interpretation is a kind of "structure neutral" interpretation (All)(cars)(garages), with further interpretation imposed by pragmatic factors in the way suggested by Bucci. Response of this type is apparently determined by the conceptual salience of 'cars in garages' i.e. the subjects' expectations about 'cars in garages'.

SUMMARY

In this chapter the ability of young children to understand universally quantified propositions of the type "All A are B." was examined, particularly in the light of Bucci's (1978) claim that, in their initial attempts to understand such statements, young children frequently encode these statements as unordered, unstructured strings (All)(A)(B). Experiment 1 involved a task similar to Bucci's "Display task" which avoided some procedural difficulties of Bucci's study. No evidence was found for a specific prediction of Bucci's theory that performance by young subjects in interpreting false propositions should be superior to that in interpreting true propositions. In fact the truth value of the statement had no effect on performance.

CHAPTER 3

VERIFICATION OF CONDITIONAL AND UNIVERSALLY QUANTIFIED SENTENCES

In experiment 1 the ability of 6 year olds and 11 year olds to perform a simple sentence/picture verification task with universally quantified propositions like "All A are B" was studied. In experiment 1 the false sentences used were always the converses of sentences which were true of the collection being considered e.g. if it was true that "All the red shapes are square." the false sentence used would be "All the square shapes are red."

Deanna Kuhn (1977) describes a task similar to the quantified verification task of experiment 1 but in which conditional rather than class statements were to be verified. Instead of simply using true and false statements Kuhn included the eight possible different types of conditional statement which can be made by combining positive and negative values of two binary propositions p and q.

The problem Kuhn used concerned different types of bug in a garden. The bugs could be big (p) or small (-p), striped (q) or black (-q). There were three different types of bug present in the garden and these were a big striped bug (pq), a small striped bug (-pq) and a small black bug (-p-q). The relationship represented in the picture (See Figure 3.1) between the binary attributes p and q could thus be said to represent the implication relationship big (p) implies striped (q) or black (-q) implies small (-p). The subject's task was to say which of the eight possible conditional sentences which can be made by combining the attributes of size and marking were true and which were false. The sentences are shown in Table 3.1 along with the truth value for a sentence in the light of different interpretations -

Figure 3.1

Picture of Insects used in Verification Task

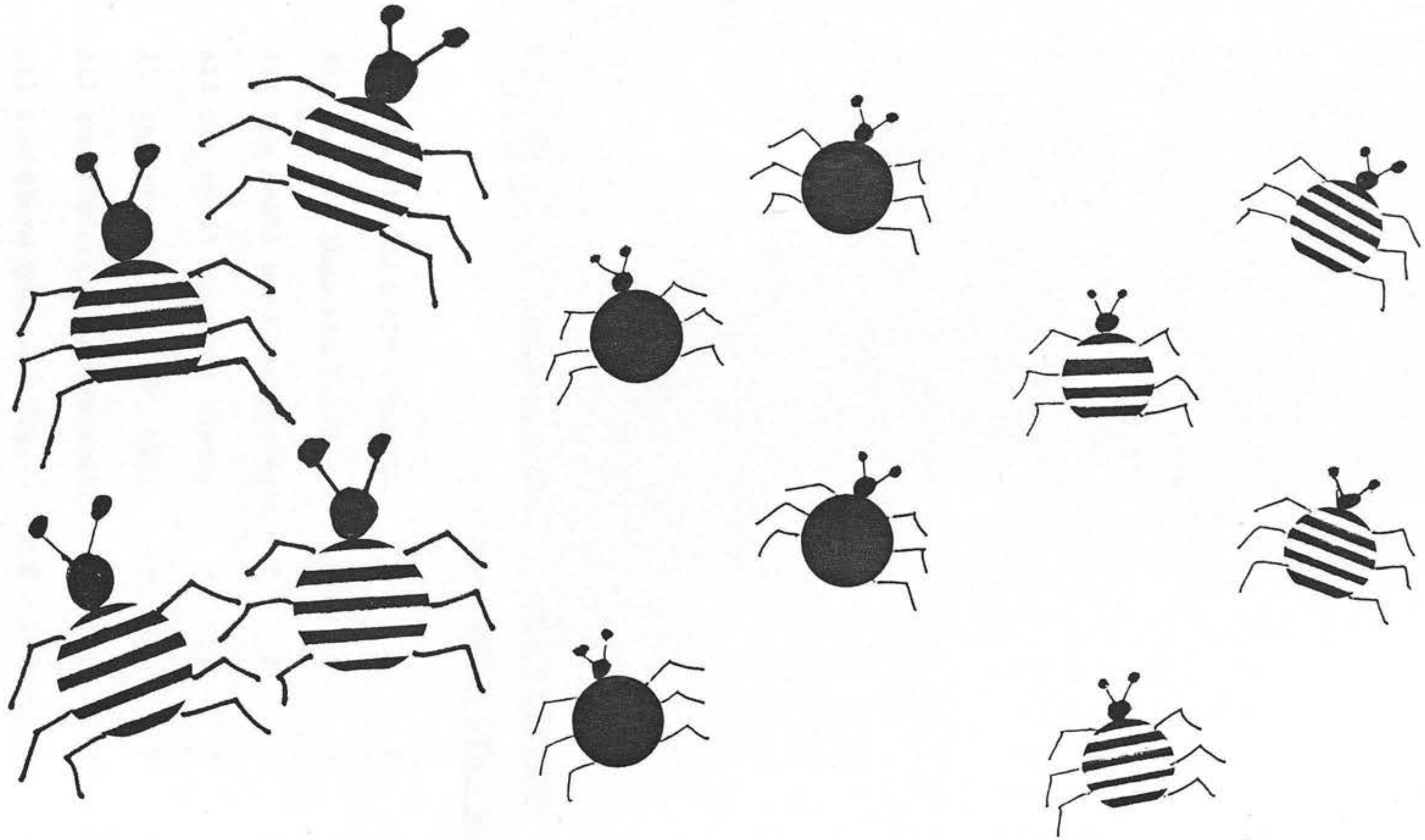


TABLE 3.1

THE EIGHT CONDITIONAL STATEMENTS USED IN THE CONDITIONAL VERIFICATION TASK WITH THE TRUTH VALUES FOR THE DIFFERENT STATEMENTS FOR DIFFERENT INTERPRETATIONS

	CODE	POSSIBLE STATEMENTS	INTERPRETATIONS		
			COND.	BICOND.	CONJ.
1)	T	If a bug is big it is striped.	T	T	T
2)	N	If a bug is big it is black.	F	F	F
3)	E	If a bug is small it is striped.	F	F	T
4)	B	If a bug is small it is black.	F	T	T
5)	B	If a bug is striped it is big.	F	T	T
6)	E	If a bug is striped it is small.	F	F	T
7)	N	If a bug is black it is big.	F	F	F
8)	T	If a bug is black it is small.	T	T	T

TABLE 3.2

THE EIGHT CLASS STATEMENTS USED IN THE CLASS VERIFICATION TASK WITH THE TRUTH VALUES FOR THE DIFFERENT STATEMENTS FOR DIFFERENT INTERPRETATIONS

	CODE	POSSIBLE STATEMENTS	INTERPRETATIONS			
			C.I.	EQU.	CON.	REV.
1)	T	All the big bugs are striped.	T	T	T	F
2)	N	All the big bugs are black.	F	F	F	F
3)	E	All the small bugs are striped.	F	F	T	F
4)	B	All the small bugs are black.	F	T	T	T
5)	B	All the striped bugs are big.	F	T	T	T
6)	E	All the striped bugs are small.	F	F	T	F
7)	N	All the black bugs are big.	F	F	T	F
8)	T	All the black bugs are small.	T	T	T	F

conditional, biconditional and conjunctive - of the sentences.

It is useful for the purposes of discussion to code the sentences. T denotes a sentence which is true under all interpretations; B denotes a sentence which is true under a biconditional and conjunctive interpretation but not under a conditional interpretation; E denotes a sentence which is true under a conjunctive but not under a conditional or a biconditional interpretation; N denotes a sentence which is false under all interpretations because no elements with the attributes mentioned in the sentence are present in the picture.

Subjects correctly interpreting the conditionals as conditionals would classify only T type sentences as true. Previous research in reasoning with conditionals has shown that children frequently misinterpret conditionals as biconditionals. Under such an interpretation T and B type sentences would be classified as true and E and N type sentences as false.

Kuhn discovered in her task that two response patterns predominated. Children over 11 years predominantly adopted the correct conditional interpretation while children up to 11 years classified T, B and E type sentences as true and N type sentences as false. In other words subjects adopting this interpretation appeared to ignore the conditionality of the sentences treating them instead as conjunctions. Sentences were classified as true if an element (i.e. a bug) with the attributes mentioned in the sentence was present in the picture.

The sharp transition from a predominantly conjunctive interpretation at 11 years to a predominantly conditional interpretation at 13 years led Kuhn to propose that the task involved conditional reasoning and that performance on the task was "significantly influenced by the development of the basic logical reasoning operations defined by Inhelder and Piaget (1958) as 'formal operations'. Kuhn claimed that the verification task with conditional sentences which she called the 'conditional reasoning task' involved the ability to comprehend a given combination of the products of the logical multiplication of two dichotomous variables p and q in the context of other possible combinations. It will be remembered from the introduction that this ability to comprehend propositional relations is one of the abilities which Inhelder and Piaget considered to be acquired at formal operations. In the traditional Piagetian hypothesis testing task described by Inhelder and Piaget (1958) (e.g. the pendulum task, the flexibility of rods task etc.) the subject was required to formulate and test hypotheses. From the child's approach to the task Piaget was able to "diagnose" the subject's operational level. Kuhn's task involves only some of the component skills involved in the hypothesis testing task and consequently is apparently simpler than these tasks. Kuhn's task does not require the subject to formulate a hypothesis for instance; rather the subject is presented with different hypotheses (conditional sentences) which he has to test (classify as true or false) with respect to a picture. The "hypothesis" or conditional sentence is true if the implication relationship which it expresses corresponds to the implication relation represented in the picture. According to Ennis's interpretation of Piaget's combinatorial logic an implication such as "If a bug is big it is striped." is true if there exists at least one big striped bug (pq), at least one not-big, striped bug (-pq) and at least one not-big, not-striped bug (-p-q) and there exist no big, not-striped bugs (p-q). Kuhn's picture represents exactly such a situation and the task is apparently a straight test of comprehension of implication. The conditional verification task is also simpler than the standard Piagetian hypothesis testing tasks in that it only includes implication relationships and not any of the other possible propositional relationships including disjunction, conjunction, equivalence (biconditional), complete affirmation etc.

To substantiate her claim that the conditional reasoning problem is a formal operational task Kuhn compared her subjects' performance on this problem with their performance on another reasoning problem said to require formal operational thinking - the isolation and testing of variables problem developed by Kuhn and Brannock (1977) and Kuhn and Angelev (1976). Kuhn found that subjects rarely showed correct reasoning on the conditional verification problem (Kuhn's conditional reasoning task) before they reached the formal level on the isolation of variables problem, a finding which she interpreted as supporting her claim that the conditional verification problem requires formal operational thinking.

It is interesting however to note the conclusions of Roberge and Flexer (1980) in a similar comparison of performance on propositional logic tasks and a control of variables task:

"the acquisition of propositional logic prior to the ability to control variables is congruent with Inhelder and Piaget's contention that the acquisition of propositional logic is a prerequisite to the complete mastery of the controlling variables procedure." (p. 9).

In both studies significant correlations were found between performance on the control of variables task and the propositional verification task for grade 6 (around 12 years) and grade 8 (around 14 years) subjects but the putative order of acquisition of the skills was different in these studies with Kuhn claiming that acquisition of the isolation of variables procedure predates conditional verification ability and Roberge and Flexer claiming that the ability to verify connected propositions is acquired prior to the ability to isolate and test variables. It appears that the contradictory conclusions concerning developmental priority are attributable to the nature of the specific items used in the tasks. Roberge and Flexer included the propositional relations biconditional and inclusive and exclusive disjunction as well as the conditional in their verification task while Kuhn only included the conditional. It is well established that biconditional and disjunctive relations are easier to comprehend and are dealt with at an earlier age than the conditional (Shine and Walsh, 1971; Suppes and Feldman, 1971; Paris, 1973; Sternberg, 1979). Consequently performance on Roberge and Flexer's propositional verification task would be predicted to be better than that on Kuhn's conditional verification task. In addition it seems that the control of variables procedure used by Roberge and Flexer was particularly difficult; only 15% and 35% respectively of grade 6 and grade 8 students respectively were classified as formal on this task. The major difficulty seemed to be in simply understanding the requirements of the task. Leaving aside the question of the order of acquisition of the ability to perform the tasks the correlations between the tasks in both studies are interpreted as indicating that similar subsets of cognitive abilities are implicated in correct solution of the two tasks.

It is not obvious that Piaget would have regarded Kuhn's "conditional reasoning task" as a test of formal operational thinking. It was pointed out in Chapter 1 that it is difficult to derive from Piaget's theory unambiguous predictions concerning the verbal deductive reasoning abilities typical of subjects at different operational stages. Piaget claimed that the major identifying feature of formal operational thought is the ability to deal with hypotheses instead of only with objects (Piaget, 1972, p. 47) although it is not obvious from his theory exactly what it is that makes hypothesis testing a formal operational ability. It is not clear whether it is the ability to formulate and test hypotheses or the ability to understand a particular propositional relationship as corresponding to a particular combination that is the hallmark of formal thinking. If it is the interpretation of empirical evidence as corresponding to a particular propositional relationship which is essentially formal operational then correct response on the conditional verification task would require formal operational thinking. However if it is the formulation and testing of hypotheses which is formal operational then the conditional verification task would not necessarily require formal operational thought.

There is a further ambiguity concerning whether the interpretation of empirical evidence in terms of propositional relationships should be regarded as a formal operational ability. On the one hand Piaget argues that the combinatorial system enables the formal operational subject to understand propositional relationships as corresponding to particular combinations of elements or "base associations". On the other hand he argues that concrete operational subjects can reason with propositions provided these correspond to "sufficiently concrete representations". If the concrete operational subject can reason with such propositions he should surely be able to say whether they are true or false with respect to "concrete representations". It is not clear whether Piaget (1972) is suggesting that the ability to verify hypotheses against concrete empirical data is a concrete or formal operational ability:

"hypotheses, not being objects, are propositions and their content consists in intrapropositional operations of classes, relations, etc., which can be verified directly; the same is true of the consequences derived from them." (p. 47)

The intrapropositional operations of classes and relations are the operations available to the subject at concrete operations and consequently this statement might be interpreted as saying that the ability to verify propositions is a concrete operational ability. However Piaget's theory is generally understood as entailing that the interpretation of empirical information as corresponding to a particular propositional relation is a formal operational ability. As we have seen the data from Kuhn's "conditional reasoning" experiment suggest that conditional verification is indeed a formal operational ability.

Kuhn does not discuss the class inclusion analogue of the conditional verification problem but it is likely that, like Piaget, she would regard it as a concrete operational ability. The class inclusion analogue of the conditional verification problem was discussed in Chapter 2. The ability to verify class statements such as "All A are B" was said to depend upon the classification operations typical of concrete operations. Presumably concrete operational subjects can verify statements like "All A are B" provided such statements correspond to "sufficiently concrete representations". In the verification task the subject is presented with an empirical array against which to verify the statement.

Although the proposal that while the class verification task is a concrete operational task the conditional verification task is a formal operational task is congruent with Piaget's theory, Inhelder and Piaget rejected the linguistic criterion for determining a subject's operational level and maintained that the subject's linguistic comprehension is constrained by the cognitive operations available to the subject rather than the linguistic expression of a relationship. This position might be understood as entailing that concrete operational subjects could understand the conditional in terms of the corresponding concrete classification operations. In fact Kuhn's data indicated that this was not the case.

Propositional theories of verification which propose that verification is a function of propositional structure (logical form) would predict no difference in the verification of class and general conditional statements since these statements have the same logical form, viz $x(Ax \rightarrow Bx)$.

The hypotheses considered in experiment 1 to explain errors in verifying universal affirmative propositions can be extended to make predictions about performance on E and N type class statements. The class logic analogues of the conditional statements in Table 3.1 are shown in Table 3.2 along with the appropriate code and response patterns for different interpretations of the class statement. The class inclusion interpretation corresponds to the conditional; equivalence to the biconditional; also shown are the "structure neutral" and the conjunctive interpretations; another possible interpretation of the class inclusion statement is the conversion interpretation under which only B type statements would be classified as true and the rest false.

Since no elements with the attributes mentioned are present in the picture for N type statements all hypotheses would predict that subjects at all ages would find it easy to classify N type statements as false.

The structure neutral hypothesis would predict that there would be no difference in performance on E and B type class statements and performance on these statements would be good at all ages. This is because any elements which do not have the attributes p.q are counterexamples to the structure neutral interpretation of "All P are Q" as "Everything is p and q". As before performance on T type statements would be predicted to be poor for younger subjects and improve significantly with age.

The F.Q.P hypothesis and the Chapmans' conversion hypothesis which made the same predictions as Bucci's S.N. hypothesis concerning T and B type statements differ from the S.N. hypothesis in predicting superior performance on E type over B type statements. According to the F.Q.P. hypothesis, in interpreting e.g. "All the small insects are striped." the subject is trying to establish whether the small insects are co-extensive with the striped insects. For E type statements both big, striped insects and small, black insects are counterexamples to this, while for the B type statements "All the small insects are black." only small, striped insects falsify the statement.

Revlin and Leirer's conversion hypothesis, which predicted no difference in correct response to T and B type statements, predicts a lower error rate on E type statements compared with B and T type statements since on conversion an E type (false) statement becomes another E type (false) statement while conversion of a T or a B type statement changes its truth value.

Experiment 2 was designed to compare the performance of subjects over the age range from the onset of concrete operational thinking to well into formal operations on the verification of universally quantified and conditional sentences. The conditional verification procedure was a replication of Kuhn's "conditional reasoning" task. From the contrasting results of the previous studies of the verification of class and conditional statements it seemed likely that the younger children at least would respond in different ways on the conditional and class verification tasks.

GENERAL PROCEDURE FOR EXPERIMENTS 2, 3, 5 AND 6

Since the same subjects performed on the tasks reported in experiments 2, 3, 5 and 6 the subjects used and the general procedure used in presenting the tasks are described here.

SUBJECTS

192 children took part in the study: 32 from each of primary grades 2, 4 and 6 from an Edinburgh primary school and 32 from each of the 1st. 3rd. and 5th. years of an Edinburgh secondary school. The mean ages and age ranges were as follows: 6 years 5 months (6 year 0 months to 6 years 10 months), 8 years 6 months (8 years 1 month to 8 years 10 months), 10 years 6 months (10 years 0 month to 10 years 11 months), 12 years 8 months (12 years 0 months to 13 years 2 months), 15 years 1 month (14 years 3 months to 15 years 9 months) 17 years 1 month (16 years 7 months to 17 years 11 months). The primary school was a feeder primary for the secondary school. Half the subjects were boys and half were girls. Primary school subjects were randomly selected by class teachers and secondary school subjects by the deputy head master as being of average to above average ability.

TASKS AND DESIGN

The sentence verification task (EXPERIMENT 2) was presented along with the picture verification task (EXPERIMENT 3), the evaluation task (EXPERIMENT 5) and the syllogistic reasoning task (EXPERIMENT 6) in order to assess how subjects performed on a variety of tasks involving the comprehension of and reasoning with class and conditional sentences. Each subject was presented with all four tasks. The results of experiments 2 and 3 are discussed in Chapter 3; experiment 5 is discussed in Chapter 5 and experiment 6 in Chapter 6.

All four tasks involved the same kinds of statement - class statements of the type "All A are B" and the logically equivalent general conditional statements of the type "If x is an A then x is a B" but the requirements of the task differed. Similar content of two different kinds - insects and shapes - was used in all four tasks. Since there were four tasks and four combinations of linguistic form and content, subjects were given a different linguistic form/content combination on each task. Since performance on the evaluation and syllogism tasks was to be compared, subjects were given statements of the same linguistic form in the evaluation and syllogism tasks. Linguistic form on the verification tasks for any subject was also the same.

Although presenting factorial combinations of content, task and presentation order would minimise the effects of order of presentation, performance on the verification task was to be compared with performance on Kuhn's conditional verification task and experiment 1 and this task was always presented prior to the picture verification task. The evaluation task was always presented before the syllogism task since Marcus and Rips (1979) have shown that, for adults at least, presentation of the evaluation task prior to the syllogism task does not affect response on the syllogism task. Consequently tasks were presented in two different orders. These were either 1) sentence verification, 2) picture verification, 3) evaluation and 4) syllogism or 3) evaluation, 4) syllogism, 1) sentence verification and 2) picture verification. In order to minimise any influence of the evaluation task on the syllogism task the content of the problem for a particular subject was changed from the evaluation task to the syllogism task. Although there are problems in using the same subjects in different tasks, analysis of presentation order of the tasks would determine whether prior presentation of the empirical verification tasks influenced performance on the reasoning tasks (evaluation and syllogism tasks) and vice versa.

PROCEDURE

Subjects were seen individually in a quiet room. The session lasted about 30 minutes for the younger children and about 20 minutes for the older children. Subjects were told that they were going to answer some problems which, although not difficult, could be a bit tricky and they would require the subject to think carefully before responding. It was hoped that this would dissuade the subjects from adopting oversimplistic strategies. Subjects were asked some preliminary questions in order to establish that they knew what true and false meant: some of the younger subjects preferred to use "right" and "wrong" and were encouraged to use these responses instead. Four practice items were presented: two sentences which were to be evaluated as true or false with respect to a picture and two sentences which were to be evaluated as true or false in the absence of a picture. The four tasks were then presented to the subject.

EXPERIMENT 2

THE SENTENCE VERIFICATION TASK

METHOD

The task was similar to Kuhn's (1977) 'conditional reasoning' task but involved the verification of either class or conditional sentences and was presented by means of two different content vehicles, either insects or shapes.

The subject was presented with a picture of three different kinds of insect (or shapes) - big striped insects (pq), small striped insects (-pq) and small black insects (-p-q): there were three insects of the three different kinds making nine in all. The relationship between the stimulus attributes, size and marking, was thus an implication relation such that big (p) implied striped (q) or black (-q) implied small (-p).

In order to provide a context for the problem a short story was read which also contained the task instructions. The story for the insect problem was as follows (that for the shape problem is in Appendix A):

"One day when Mr. Jones was in his garden he noticed lots of insects. He watched carefully and saw three different kinds of insect. This is what they looked like...(PICTURE PRESENTED)...Later he was telling a friend about the insects. Which of the following sentences that Mr. Jones might say about the insects he saw in his garden are true and which are false?"

The subject was then presented with the eight sentences which represented all the possible implication relations between size and marking. The sentences were the same as those in Tables 3.1 and 3.2 except that the word 'insects' was substituted for the word 'bugs'. Each sentence was printed in large letters on separate cards which were presented to the subject one at a time in random order. The subject's task was to say, for each card, whether the sentence on the card was true or false with respect to the picture. The subject was encouraged to read each sentence aloud and to examine the picture carefully before responding. Half of the subjects were given class statements like "In my garden all the big insects are striped " and half were given the logically equivalent conditional statements like "In my garden if an insect is big it is striped.". Linguistic form and content were combined factorially and subjects were assigned to groups in the way described in the general procedure.

RESULTS

In the original analysis content (shapes/insects) was included as a factor and the data were dichotomous. Lunney (1970) has shown that analysis of variance can be validly applied to dichotomous data provided there are sufficient degrees of freedom for error (at least 20). Since this criterion was met an analysis of variance was applied to the data.

There was no main effect of content and no interaction of content with any other factor was significant. Consequently the analysis with content as a factor is not reported and further analysis was performed excluding this factor.

Table 3.3 shows mean percent correct response to the different statement types, T, N, E and B for the two different linguistic forms, class and conditional, for subjects at different grades and this is

TABLE 3.3

MEAN PERCENT CORRECT RESPONSE ON DIFFERENT STATEMENT TYPES (T, B, E, N) FOR DIFFERENT LINGUISTIC FORMS (CLASS AND CONDITIONAL) ACROSS GRADE

	T			B			E			N			TOT		
	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT
P2	94	88	91	63	22	42	72	22	47	94	100	97	81	58	70
P4	78	91	84	56	22	39	72	28	50	100	98	99	77	60	68
P6	88	78	83	72	19	45	88	31	59	94	97	95	86	56	71
S1	84	78	81	100	53	76	100	72	86	100	100	100	96	76	86
S3	97	81	89	84	78	81	97	91	94	100	100	100	95	88	91
S5	88	84	86	91	91	91	100	81	90	100	100	100	88	89	88
TOT	88	83	85	78	48	63	88	54	71	98	99	98	88	71	80

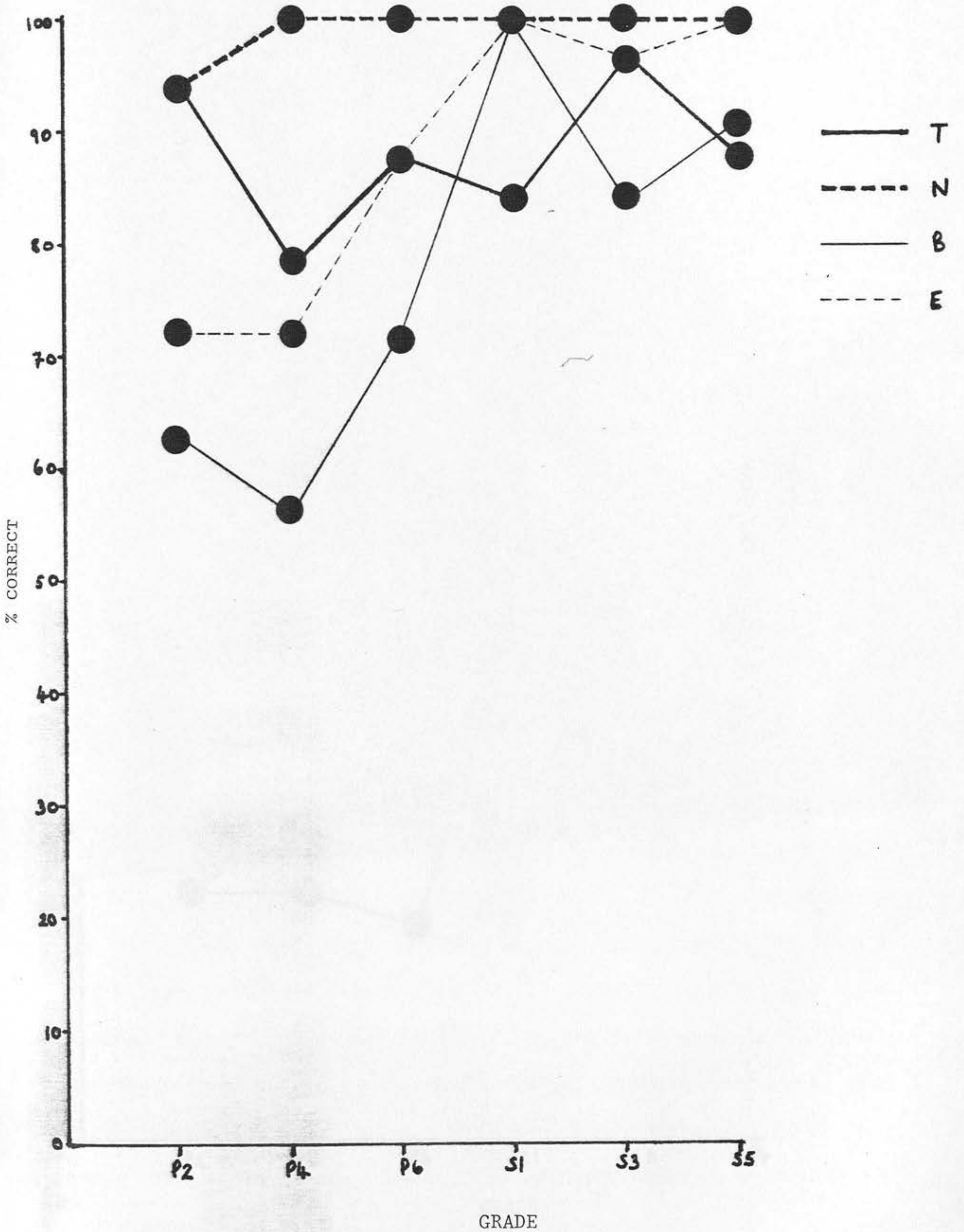


FIGURE 3:2a

PERCENT CORRECT RESPONSES TO THE DIFFERENT STATEMENT TYPES (T,N,E,B) FOR CLASS STATEMENTS PLOTTED AGAINST GRADE

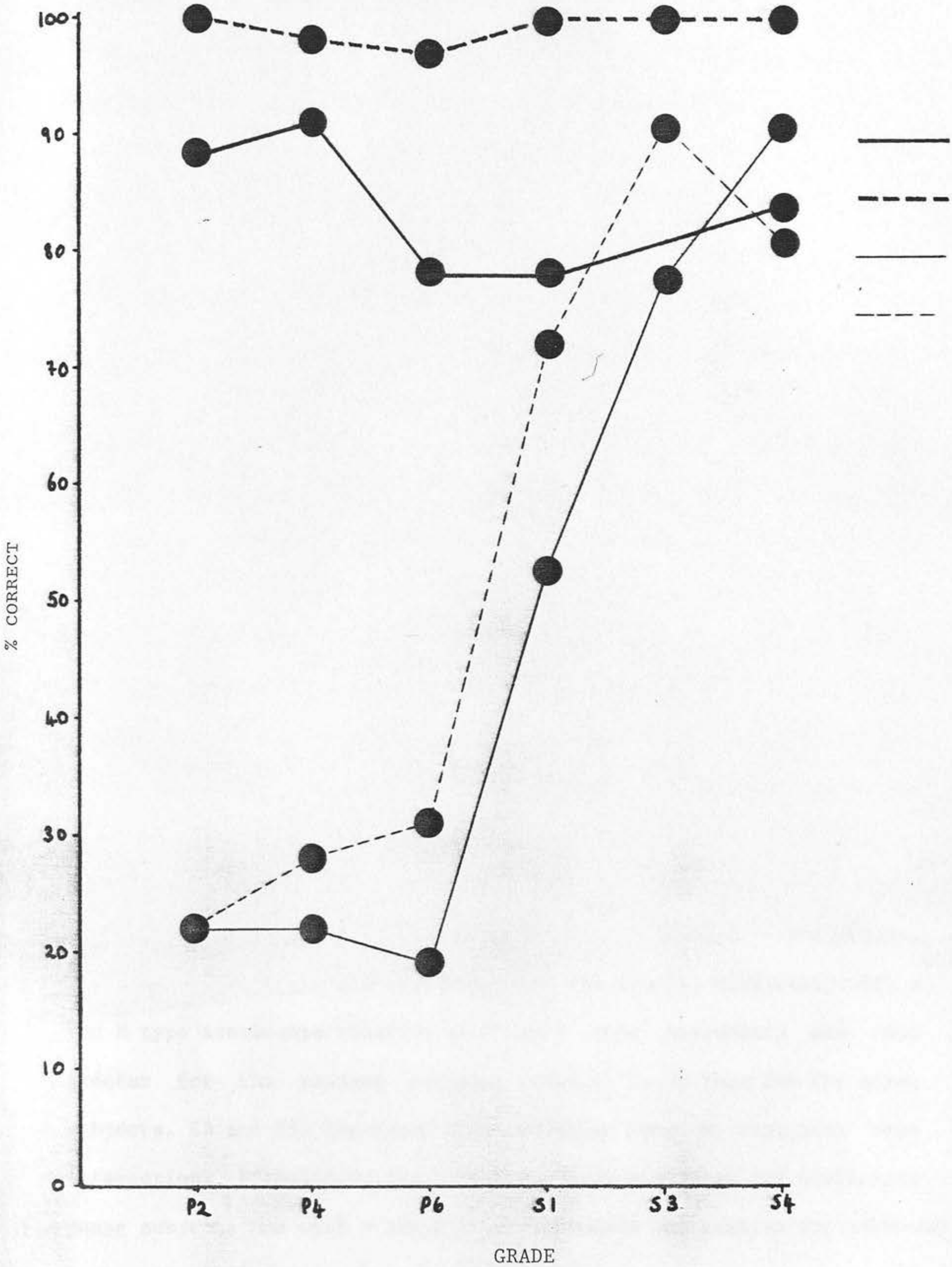


FIGURE 3:2b

PERCENT CORRECT RESPONSES TO THE DIFFERENT STATEMENT TYPES (T,N,E,B) FOR CONDITIONAL STATEMENTS PLOTTED AGAINST GRADE represented graphically in Figures 3.2a and 3.2b. The results of a 6 (grade) X 2 (linguistic form) X 4 (statement type) analysis of variance with grade and linguistic form as between subjects factors and statement type as a within subjects factor are shown in Table 3.4. All factors and interactions were significant. The significant effect of grade, $F(5,180)=19.05$, $p<0.001$, was mainly attributable to the superior performance of secondary compared with primary subjects. The significant main effect of linguistic form, $F(1,180)=63.66$, $p<0.001$, showed that overall class statements (88% correct) were responded to better than conditional statements (71% correct). The significant effect of statement type, $F(3,540)=73.24$, $p < 0.001$, showed that performance on some statement types is better than that on others: specifically the order of difficulty with the easiest first was N T E B. The grade X linguistic form interaction, $F(1,180)=3.09$, $p<0.05$, indicated that the effect of linguistic form was significantly greater at some grades than others. The highly significant linguistic form X statement type interaction, $F(15,540)=22.65$, $p<0.001$, reflected the fact that the superior performance on class statements was almost entirely attributable to the superior performance on E and B type class statements. The grade by statement type interaction, $F(15,540)=7.73$, $p<0.001$, indicated that the general difficulty with B and E type statements relative to N and T type statements was much greater for the younger subjects i.e. up to S1 than for the older subjects, S3 and S5. The grade X linguistic form X statement type interaction, $F(15,540)=2.52$, $p<0.005$, indicated that the difficulty young subjects had with B and E type statements was greater for conditional statements than for class statements.

To test the simple main effects of linguistic form at grade(i), the F ratio has the form F=MS(b at a(i))/MS(error between subjects) (see

TABLE 3.4

ANALYSIS OF VARIANCE FOR GRADE, STATEMENT TYPE AND LINGUISTIC FORM

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
MEAN	1938.021	1	1983.021	5606.73	0.001
GRADE	32.917	5	6.583	19.05	0.001
LINGUISTIC FORM	22.005	1	22.005	63.66	0.001
GL	5.339	5	1.068	3.09	0.05
ERROR	62.219	180	0.346		
STATEMENT TYPE	58.094	3	19.365	73.24	0.001
SG	30.656	15	2.044	7.73	0.001
SL	17.964	3	5.988	22.65	0.001
SGL	10.005	15	0.667	2.52	0.005
ERROR	142.781	540	0.264		

TABLE 3.4a

SUMMARY TABLE FOR SIMPLE EFFECTS OF LINGUISTIC FORM AT GRADE(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	6.570	1	6.570	18.99	0.001
P4	3.780	1	3.780	10.92	0.001
P6	10.695	1	10.695	30.91	0.001
S1	5.281	1	5.281	15.26	0.001
S3	0.633	1	0.633	1.83	ns
S5	0.383	1	0.383	1.11	ns
ERROR	62.219	180	0.346		

TABLE 3.4b

SUMMARY TABLE FOR SIMPLE LINGUISTIC FORM X STATEMENT TYPE (LS) INTERACTION AT GRADE(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	6.961	3	2.320	8.79	0.001
P4	6.656	3	2.219	8.41	0.001
P6	8.773	3	2.924	11.08	0.001
S1	4.406	3	1.469	5.56	0.001
S3	0.398	3	0.133	0.50	ns
S5	0.773	3	0.258	0.98	ns
ERROR	142.781	540	0.264		

footnote 3.1). The results of this analysis are shown in Table 3.4a. Significant F ratios, calculated against degrees of freedom (1, 180), were found at all primary grades and at first year secondary but not at third and fifth year secondary.

To test the simple interaction of linguistic form by statement type the F ratio has the form F=MS(bc at a(i))/MS(error at abc). Table 3.4b shows that the interaction was significant at all primary grades and at first year secondary but not at third or fifth year secondary. The significant linguistic form by statement type interaction indicates that different patterns of response were given on the class and conditional verification problems at least for subjects up to first year secondary. Further analysis was performed separately on the data from class and conditional statements.

VERIFICATION OF CLASS STATEMENTS

RESULTS

A 6 (grade) X 4 (statement type) analysis of variance with grade as a between subjects factor and statement type as a within subjects factor was carried out on correct response on class statements as a preliminary to analysing simple effects (see Table 3.5). Both factors, grade $F(5,90)=5.10$, $p<0.001$, and statement type $F(3,270)=14.13$, $p<0.001$, and the grade by statement type interaction, $F(15,270)=2.76$, $p<0.005$, were significant.

Tests of simple interactions were carried out on adjacent grades (Bruning and Kintz, 1968, p.120). Only the comparison between P6 and S1 was significant, $F(3,270)=2.95$, $p<0.05$ (see Table 3.5a). However it seems likely that this was due to the exceptional performance by first

TABLE 3.5

ANALYSIS OF VARIANCE FOR CLASS STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
TOTAL	106.52	383			
BETWEEN SUBS	40.23	95			
GRADE	8.87	5	1.774	5.10	0.001
ERROR	31.36	90	0.348		
WITHIN SUBS	66.29	288			
STATEMENT TYPE	7.93	3	2.643	14.13	0.001
GS	7.74	15	0.516	2.76	0.005
ERROR	50.62	270	0.187		

TABLE 3.5a

TESTS OF SIMPLE INTERACTIONS FOR GRADE X STATEMENT TYPE INTERACTIONS FOR CLASS STATEMENTS: F VALUES AND SIGNIFICANCE LEVELS FOR BETWEEN GRADE COMPARISONS

	P2	P4	P6	S1	S3	S5
P2	-					
P4	0.279 ns	-				
P6	1.631 ns	1.824 ns	-			
S1	7.684 $p<0.001$	6.947 $p<0.001$	2.952 $p<0.05$	-		
S3	2.059 ns	2.715 $p<0.05$	0.107 ns	2.283 ns	-	
S5	4.955 $p<0.005$	4.378 $p<0.005$	1.112 ns	0.454 ns	0.781 ns	-

TABLE 3.5b

SUMMARY TABLE OF SIMPLE MAIN EFFECT OF STATEMENT TYPE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	4.797	3	1.599	8.55	0.001
P4	6.312	3	2.104	11.25	0.001
P6	1.672	3	0.557	2.98	0.05
S1	1.172	3	0.391	2.09	0.1
S3	0.922	3	0.307	1.64	ns
S5	0.797	3	0.266	1.42	ns
ERROR	50.62	270	0.187		

TABLE 3.5c

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF GRADE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
T	1.428	5	0.286	1.25	ns
N	0.333	5	0.067	0.28	ns
E	5.678	5	1.136	4.98	0.001
B	9.178	5	1.836	8.05	0.001
ERROR	81.980	360	0.228		

TABLE 3.5a

NEWMAN KEULS COMPARISONS ON STATEMENT TYPE MEANS FOR CLASS STATEMENTS

STATEMENT TYPE	B	T/E	N	r	sq(0.99)(r,270)
ORDERED MEANS	1.552	1.760	1.958		
B 1.552	-	0.208 **	0.406 **	3	0.181
T/E 1.760		-	0.198 **	2	0.160

year secondary subjects who performed even better than third and fifth year secondary subjects on this task. Non-adjacent comparisons revealed no significant primary/primary or secondary/secondary comparisons and the comparisons between P6 and S3 and between P6 and S5 subjects were not significant although all other primary/secondary comparisons were significant.

To test the simple main effect of statement type the F ratio has the form (Winer, 1971, p.529), F=MS(statement type at grade(i))/MS(error between subjects). The effect of statement type was significant for all primary grades but for no secondary grades (see Table 3.5b).

The F ratio for the test on simple main effects of grade has the form F=MS(grade at statement type(i))/MS(within cell) (Winer, 1971, p.529) and was calculated against df=(5,329) (See Footnote 3.2). Significant F ratios were found for B type, $F(5,329)=8.05$, $p<0.01$, and E type, $F(5,329)=4.98$, $p<0.001$, statements but not for T or N type statements (See Table 3.5c).

Newman Keuls comparisons on the means for the different statement types collapsed across grade were carried out (see Table 3.5d). Means for T and E type statements were the same but all other comparisons were significant. Although the means for T and E type statements collapsed across grade were similar the significant main effect of grade for E type but not for T type statements reflects the fact that performance on E type statements was poorer for primary subjects.

CLASSIFICATION OF RESPONSE PATTERNS

The data were also examined in terms of the individual subjects' patterns of response to all eight statements in order to discover whether subjects adopted consistent interpretations of the quantified statements and, if so, whether the interpretation adopted changed with age. The systematic interpretations considered were the conjunctive , equivalence (biconditional) and class inclusion (conditional) interpretations (see Table 3.6). A modified classification of response pattern data was also carried out in which subjects responding to a particular interpretation except for one error were classified under that response pattern. The modified classification will be discussed only where it differs from the standard classification.

From Table 3.6 it can be seen that the only systematic response pattern adopted with any regularity was the correct one - the class inclusion pattern: overall 54% of response patterns were of this kind. Some subjects at all grades responded according to this pattern and although there was an increase in the number of correct response patterns with increasing age there were no significant comparisons of response pattern distributions even for the P2/S1 comparison. An analysis of response consistency across grade showed that only for the modified response pattern classification was there a significant increase in consistency of response with increase in age $\chi^2(5)=27.08$, $p<0.0001$.

Very few response patterns were conjunctive -only 3% - and there were no equivalence interpretations of the class inclusion statement. In short errors made in verifying class statements were apparently not due to any consistent misinterpretation of the class inclusion statement.

DISCUSSION

TABLE 3.6

CLASSIFICATION OF RESPONSE PATTERNS

CLASS

GRADE	CONJ	BICOND	COND	INCONSISTENT
P2	1	0	6	9
P4	2	0	5	9
P6	0	0	7	9
S1	0	0	12	4
S3	0	0	10	6
S5	0	0	12	4
TOTAL	3	0	52	41

CONDITIONAL

GRADE	CONJ	BICOND	COND	INCONSISTENT
P2	8	0	0	8
P4	8	1	0	7
P6	7	2	0	7
S1	3	1	2	10
S3	1	0	8	7
S5	0	0	9	7
TOTAL	27	4	19	46

TOTAL

GRADE	CONJ	BICOND	COND	INCONSISTENT
P2	9	0	6	16
P4	10	1	5	16
P6	7	2	7	16
S1	3	1	14	14
S3	1	0	18	14
S5	0	0	21	11
TOTAL	30	4	71	87

The results of experiment 2 showed that primary school children perform well on the verification task with universally quantified sentences with even P2 children (mean age 6.5 years) correct on 81% of items overall. The response pattern classification indicated that the errors which the primary children did make were not attributable to any particular systematic misinterpretation of the class statements although the statement type analysis showed that more errors are made on some statement types than others: 98% of N type statements, 88% of both T and E type statements and 78% of B type statements were verified correctly.

If the absence of a simple main effect of statement type is taken as an index of mature performance on the task then only at S1 do subjects perform maturely. However the nonsignificance of the P6/S3 and P6/S5 comparisons on the tests of simple interactions for the statement type analysis indicated that P6 subjects (average age 10.5 years) perform very well on the task.

The statement type analysis showed that the subjects made the highest number of errors on B type statements and the significance of the simple main effects analysis of grade for only B and E type statements showed that the improvement in performance with increasing age was attributable only to improvement with these statement types. Although the mean correct response to T type statements was exactly the same as for E type statements the absence of a simple main effect of grade for T type statements showed that errors on T type statements were distributed across all age groups whereas errors on E type statements were made mainly by primary school subjects.

While the results of experiment 1, in which subjects responded correctly to the same number of true (T type) and false (B type) statements, did not support Bucci's structure neutral hypothesis, the results of experiment 2 are exactly the opposite of those predicted by Bucci. The S.N. hypothesis would have predicted no significant difference in correct response to false items (B and E type statements) with increasing age but a significant increase in correct response to true items (T type) with increasing age. In experiment 2 there was no significant increase in correct response to T type statements but there was a significant increase in correct response to both B and E type statements. In addition the response pattern analysis showed that only 3 subjects over all responded according to the structure neutral response pattern.

Both Piaget's F.Q.P. hypothesis and the Chapman's conversion hypothesis made similar (incorrect) predictions to those of Bucci concerning response on T and B type statements. Both hypotheses correctly predicted however the superior performance on E type over B type items since two different kinds of element from the picture falsified an E type statement but only one kind of element falsified a B type statement.

While the results of experiment 1 were compatible with Revlin and Leirer's conversion hypothesis the superior performance on T type over B type statements in experiment 2 was not. Revlin and Leirer's conversion hypothesis, along with the Chapmans' and the F.Q.P. hypotheses, did however predict the superior performance on E over B type statements.

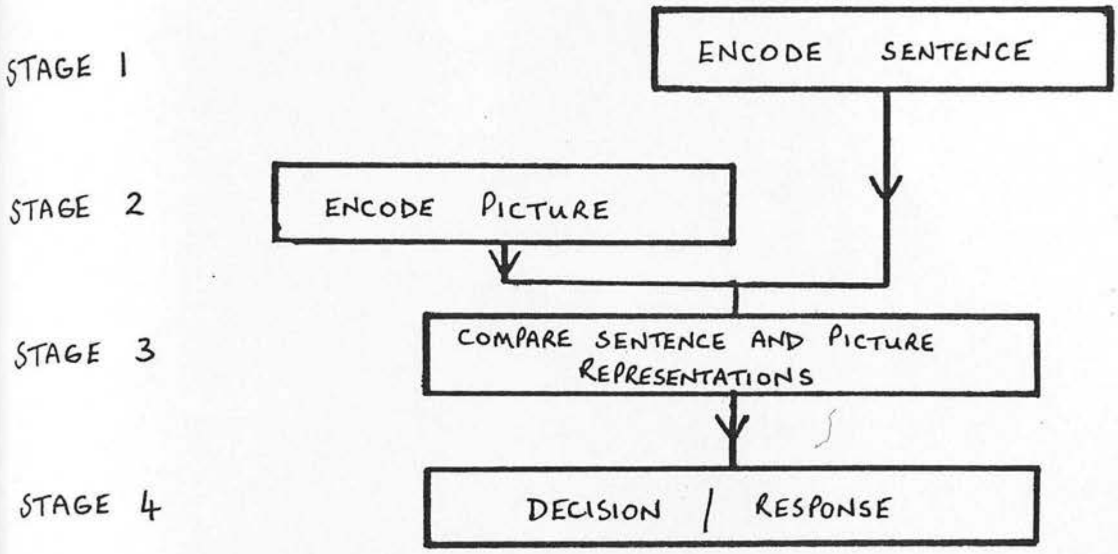
Just (1974) describes a task very similar to the class verification task of experiment 2 in which adults rather than children were required to verify quantified statements with respect to pictures depicting different kinds of relationship between the subject and predicate of a sentence: the relations included subset, superset, disjoint, overlap and identity. Just considered two models, the comparison model and the computation model, which correctly predicted some, but not all, response latencies and error rates on this sentence/picture verification task. It is of interest to consider whether either of these models could explain the errors in experiments 1 and 2 and, more generally, whether these models are appropriate as models of children's thinking.

Slightly different versions of the general comparison model were developed by Trabasso et al (1971), Clark and Chase (1972) and Carpenter and Just (1973) but they all comprised four basic stages as shown in Figure 3.3a.

At STAGE 1 the information from the sentence is encoded. The representation of information is held to be in a propositional format. In Just's notation, for instance, the sentence "All A are B" would be represented as Aff(All(Aff(A is a B))).

At STAGE 2 the information from the picture is encoded. The information from the picture is represented in a similar, i.e. propositional, format to the sentence and when the picture follows the sentence the information encoded from the picture depends on the sentence representation. A picture depicting a subset relation between A and B would have the same representation as the class inclusion statement "All A are B" viz Aff(All(Aff(A is a B))); a superset relation would be represented as Neg(All(Aff(A is a B))); an overlap relation would be given the same representation as a superset relation; a picture depicting a disjoint relation between A and B would be represented as Aff(All(Neg(A is a B))).

FIGURE 3.3a
4-STAGE MODEL OF SENTENCE/PICTURE COMPARISON



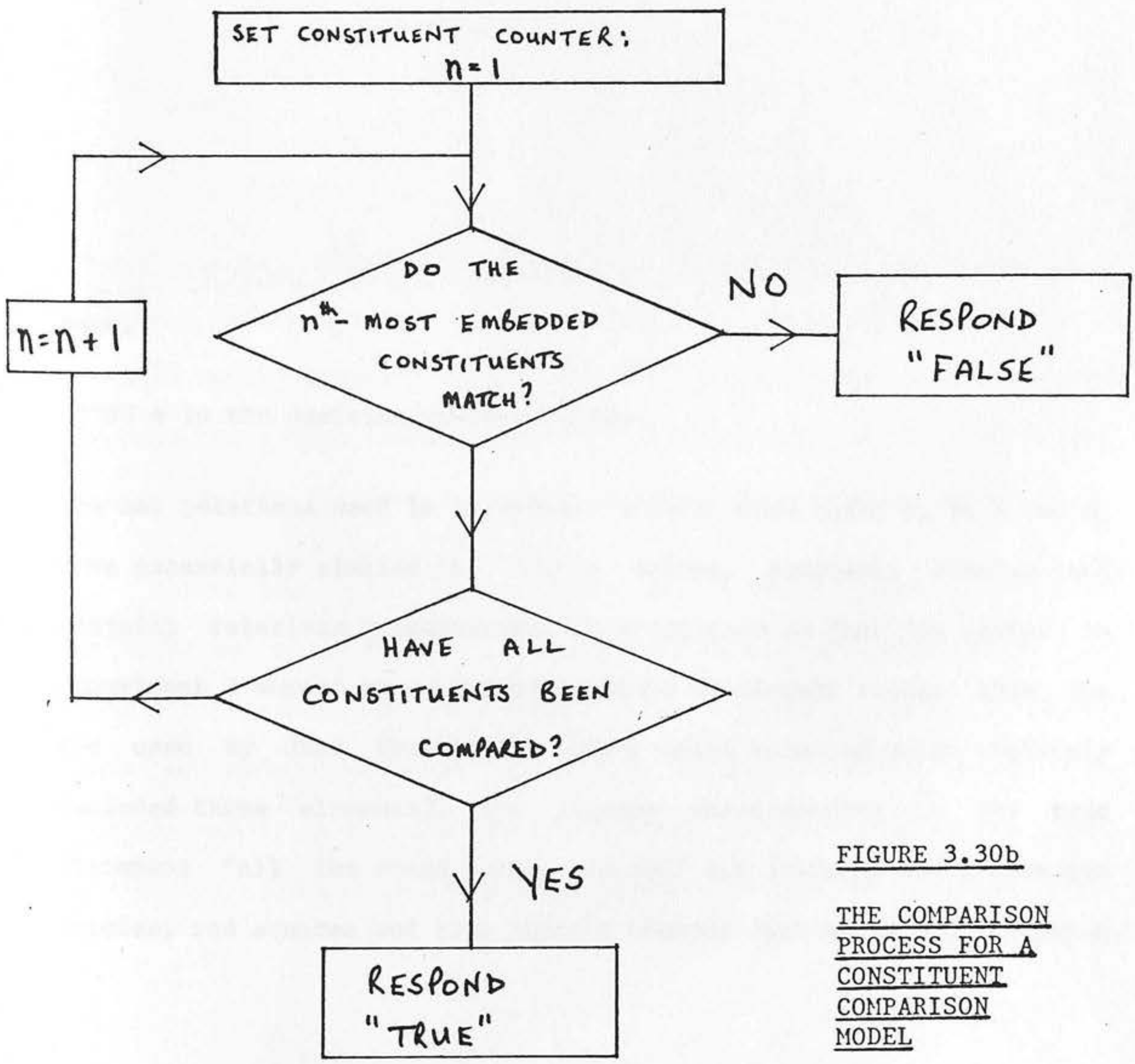


FIGURE 3.30b

THE COMPARISON PROCESS FOR A CONSTITUENT COMPARISON MODEL

STAGE 3 is the comparison stage, the central component of the comparison model. Figure 3.3b shows the sequence of operations in the comparison stage. Sentence and picture representations are in a similar format and are compared by matching corresponding constituents of the two representations starting with the most embedded constituents and proceeding until a mismatch is found or until all constituents have been compared. In verifying "All A are B" against a picture depicting a superset or an overlap relation between A and B all constituents would have to be compared since a mismatch only occurs on the outermost constituents. Similarly in verifying a subset relation all constituents would have to be compared in order to establish that there was no mismatch. A mismatch on the innermost constituents would occur in verifying the class inclusion statement against a disjoint picture. Since difficulty of verification, as measured by response latencies or error rates, is dependent upon the number of comparisons required, subset, superset and overlap verifications would be predicted to have the same error rates while disjoint verifications would be predicted to have a very low error rate.

STAGE 4 is the decision/response stage.

The set relations used in experiment 2 which were coded T, B, E and N, were essentially similar to Just's subset, superset, overlap and disjoint relations respectively. They differed in that the picture in experiment 2 showed three different kinds of element rather than the two used by Just (except in his overlap relation which obviously included three elements). The picture corresponding to the true statement "All the round shapes are red" for instance would show red circles, red squares and blue squares whereas Just's subset relation would simply show red circles and red squares. However the representation of information from the picture is held to depend upon the representation of information from the sentence and this would concern only the subject and predicate terms, which in the case of the sentence above would be round shapes and red shapes.

The constituent comparison model would predict few errors on N type statements (disjoint relationships) since innermost constituents mismatch and no further processing is required. Higher error rates would be predicted for T, E, and B type statements: since these require similar amounts of processing the error rates on these three statements would be predicted to be the same although it seems reasonable to suppose that more errors on B-type and E-type statements would occur than on T-type statements because subjects might not be sufficiently rigorous in comparing all constituents. If subjects failed to compare all constituents an incorrect 'true' response would be given for B-type and E-type statements while a similar lack of rigour in comparison of constituents on the true (T-type) statement would still result in a correct 'true' response. The comparison model does not account for the higher levels of correct response on T type compared with B type statements nor for the higher levels of correct response on E type compared with B type statements since B and E type statements have the same representation in the comparison model.

The attraction of the comparison model of sentence-picture verification is that the kinds of operations proposed are the same kinds of operations as those proposed to explain results in other kinds of deductive reasoning tasks, linguistic-based tasks, e.g. semantic memory tasks, and general sentence comprehension tasks. There is a lot of evidence to support the assumption that verification tasks are essentially linguistic tasks in which verification times depend upon the extent of propositional embedding in a sentence. This has been shown to be true with sentences expressing many different types of semantic relation, including quantifiers (Just and Carpenter, 1971; Just,1974), passives (Gough, 1966), and counterfactuals (Carpenter, 1973). In addition sentence verification latencies have been found to correlate relatively well with measures of general verbal comprehension (Hunt, 1981).

Despite the success of the constituent comparison model in explaining the results of many verification studies with adult subjects Just found that the model could not account for some of the effects observed in his experiment. He found for instance that factors relating to the presentation of the task differentially affected response times and error rates in verifying superset (B type)and subset (T type) relations. Specifically he found that error rates and latencies on subset and superset relations could be changed by altering the stimulus and response probabilities. When the probability of a subset relation (i.e. a true statement) was increased the latency and error rate associated with a true response was increased while the latency and error rate associated with the superset relation (a false statement) was decreased. Just concedes that the comparison model cannot account for the changes in difficulty on the subset and superset relations brought about by changing stimulus and response probabilities. Given that this type of presentation factor influences the responses of adults on sentence verification tasks it seems likely that similar effects will be found with children's responses on verification tasks.

One of the differences between experiment 1 and experiment 2 was just such a change in stimulus and response probabilities: this was brought about by the inclusion in experiment 2 of the N and E type statements. In experiment 1 both stimulus and response probabilities were balanced: only two types of statement were used and the presentation of these statement types was balanced. Since the correct response to one of the statements was true and to the other false the response probability was also balanced. In experiment 2 four different types of statement were used and these were presented with an equal probability. Although the stimulus probability was balanced the response probability was not and a 'false' response was three times as likely as a 'true' response. In experiment 2 the decrease in probability of a true response was accompanied by an increase in correct response to T type statements (94% correct) relative to experiment 1 (70% correct) for primary 2 subjects (P2 subjects were chosen for a comparison with experiment 1). This contrasts with Just's results since he found that increasing the probability of a true (subset) statement increased correct response on that statement. It seems likely that these different effects of changing the stimulus and response probabilities are due to the ways in which the probabilities were changed. In experiment 2 new kinds of sentence/picture relationships were introduced while in Just's task the same relations were presented more or less frequently. It is likely that in experiment 2 where four different statement types were used the subjects may contrast the N type statements, for which a mismatch is found at an early stage in processing, with all other statement types. If a preliminary response was based upon a contrast between statements for which an early mismatch was found (false) and those for which no early mismatch was found (true) an increase in correct true responses to T type statements and incorrect true responses on B type statements would be expected relative to experiment 1. In experiment 3 only 63% of B type responses for P2 subjects were correct compared with 70% in experiment 1.

Neither the comparison model nor any of the other models considered can account for differential changes in error rates on particular kinds of statement brought about by changes in presentation of the task. Any theory which was to explain the superior performance on B type statements over T type statements found by Bucci, the similar error rates on B and T type statements found in experiment 1 and the superior performance on T type over B type statements in experiment 2 would have to be very flexible.

Another model which Just (1974) extended to the domain of verification of quantified sentences was the computation model. This model is possibly more appropriate as a model of verification for younger children since it does not presuppose, as the comparison model does, that information from sentence and picture is encoded in a propositional format.

The computation model has 2 stages (Figure 3.4). At stage 1 a decision is made as to whether the subject (S) intersects the predicate (P) i.e. whether the subject and predicate have any elements in common. A NO response at this stage (an N type statement) would be relatively fast and error free. The low error rate on N type statements in experiment 2 accords with this prediction. A YES response at this stage would indicate that further processing is required to determine the exact nature of the relationship between S and P. At stage 2 a decision is made concerning whether S is a subset of P. If S is a subset of P (i.e. a T type statement) a true response would be made, otherwise if S is not a subset of P (either a B or an E type

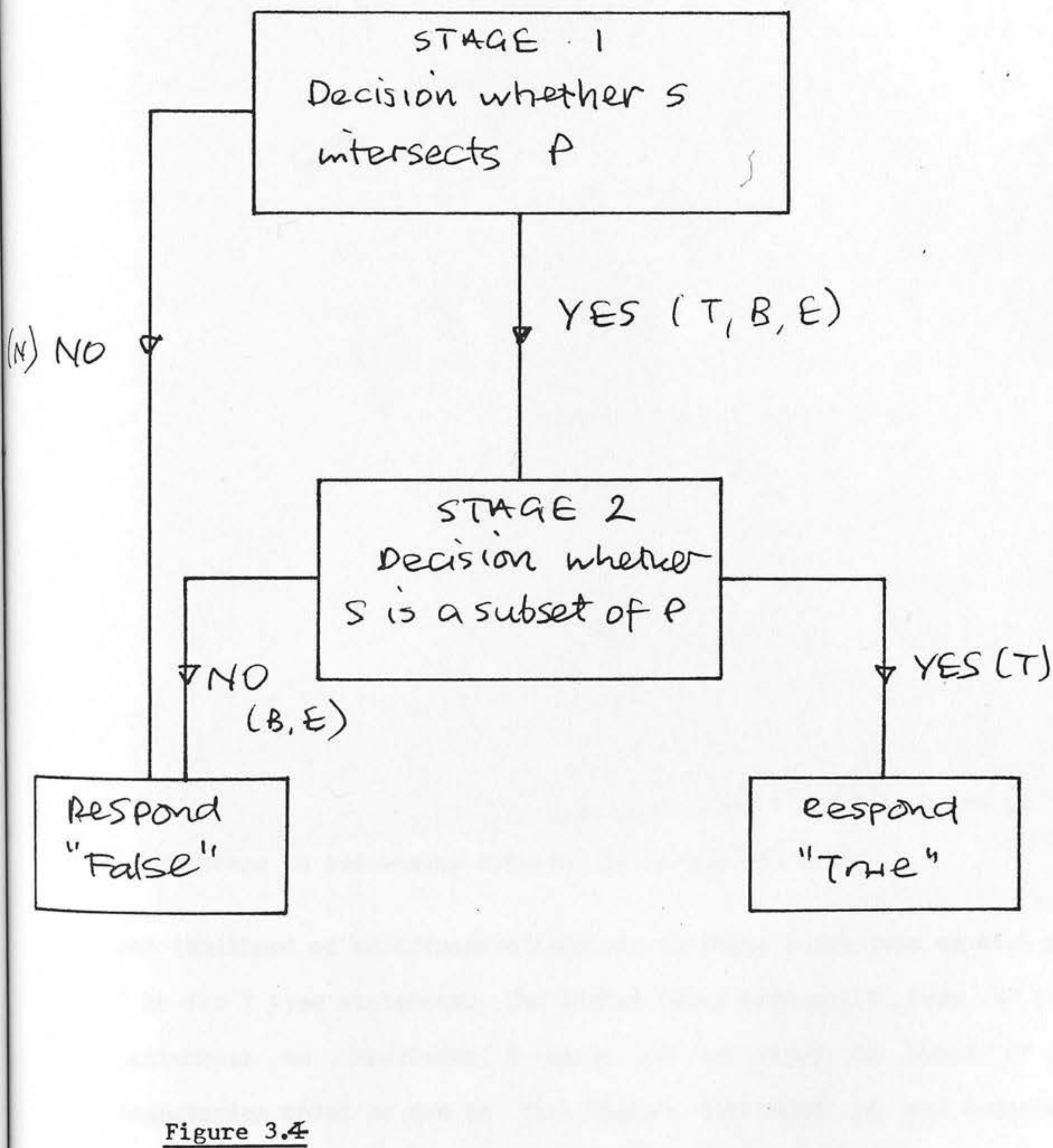


Figure 3.4

Computation Model of Verification

statement) a false response would be made. Decisions requiring stage 2 processing (T, E and B type statements) would be predicted to take longer or have more errors than those not requiring stage 2 processing (N type statements). So far the model does not make very interesting predictions. However if it is assumed that subjects sometimes fail to execute stage 2 processing, responding prematurely after stage 1, it is evident that more errors would be predicted on B and E type statements compared with T type statements since the response given after stage 1 on these statements would be an incorrect 'true' response. Premature response after stage 1 on a T type statement would still result in a correct 'true' response. The primary children in experiment 2 did indeed make more errors on B type (64% correct averaged over P2, P4 and P6) and E type (77% correct) statements than on T type statements (87% correct). In experiment 1 stage 1 is unnecessary since only B and T type statements were included. S always intersected P and the greater number of errors on B type statements relative to T type statements caused by premature response would not be predicted and was not found (70% correct on both B and T type statements). The computation model can explain the effects of changing stimulus/response probabilities in experiments 1 and 2 in terms of the extra stage in processing required in experiment 2.

The likelihood of an erroneous response at stage 2 was just as high for B as for T type statements. The higher error rate on B over E type statements on experiment 2 could be explained in terms of the computation model as due to the higher likelihood of an incorrect assessment of whether S is a subset of P for B type statements than E type statements.

While Bucci's structure neutral hypothesis was not supported by the experimental evidence of either experiments 1 or 2 Bucci's own data indicate that certain subjects in certain situations do adopt a sentence representation in which the relation between S and P is not clearly represented. In a representation of the type Bucci proposes, (All)(S)(P), the appropriate decision to make is not whether S is a subset of P, or whether S intersects P, or even whether S is equivalent to P but whether all relevant elements are both S and P.

It is interesting to consider a proposal of Revlin and Leirer's (1979) in attemping to explain the diverse results of the studies on the verification of quantified sentences. They suggest that in comprehending a sentence such as "All A are B" the subject constructs a "meaning stack" consisting of increasingly complex levels of meaning built up by progressive extraction of elementary features of the sentence from orthographic through phonological and syntactic features. Successively higher levels of the stack correspond to more recently derived meanings. Figure 3.5 shows the meaning stack which Revlin and Leirer proposed for the universally quantified sentence "All A are B". According to Revlin and Leirer the topmost level of the stack is the converted meaning of the universally quantified statement. Revlin and Leirer's conversion hypothesis seems to be quite implausible since, as the meaning stack shows, the correct inclusion interpretation of the sentence has to be derived before the converted inclusion interpretation can be derived. It is not this aspect of their theory which is of interest here however: it is their account of comprehension in terms of a meaning stack and especially their claim that progressively higher levels of the stack reflect not just a temporal processing sequence but also a developmental sequence. For the younger child the topmost level in his meaning stack might be similar to that described by Bucci as the structure neutral

FIGURE 3.5

MEANING STACK FOR UNIVERSALLY QUANTIFIED STATEMENT "ALL A ARE B"

STACK LEVEL	SENTENCE CODE
N	QUANTIFIER(PREDICATE(PREDICATE IS A SUBJECT))
N-1	QUANTIFIER(SUBJECT IS A PREDICATE)
.	
.	
.	
.	
.	
.	
K	(QUANTIFIER)(SUBJECT)(PREDICATE)
K-1	(QUANTIFIER)(SUBJECT)
K-2	(QUANTIFIER)
.	
.	
.	
.	
.	
.	
K-i	PHONOLOGICAL FEATURES
.	
.	
.	
.	
.	
.	
K-j	ORTHOGRAPHIC FEATURES

interpretation. The topmost level would become increasingly complex as the subject was able to extract a meaning which was more logically correct. It seems reasonable to propose that the same subject may access different levels of the meaning stack depending on the nature of the task, the content of the task, the effort the subject puts into solving the task and other factors. It seems likely for instance that a higher level in the stack would be accessed in a task involving concrete content since the concrete content would provide concrete referents either in the sense of physically present referents for the linguistic terms or in the sense of concrete/familiar linguistic terms. It also seems plausible that in certain situations pragmatic factors associated with the task will induce a lower level representation of the sentence. In one task - a "Block Building" task - Bucci found that even adults could be induced to adopt a "structure neutral" interpretation of a universally quantified statement and would use only big yellow bricks when instructed to make a building in which "All the big bricks are yellow.".

Stedmon (1982) drew a distinction between tasks which involve an understanding of the intensional aspects of quantification and those which involve an understanding of the extensional aspects of quantification. She argues that many class inclusion problems involve comprehension of an intensional logic relationship and that this kind of relationship is understood at an earlier age than an extensional logic relationship. An example of the former kind of problem would be asking the child as in the standard Piagetian class inclusion problem "Are all roses flowers?" or "Are there more flowers or more primroses?". Although the latter type of problem could be solved by examining the relevant set of flowers, the correct response can be derived in the absence of an extensional array of flowers by a subject who understands the intensional class inclusion relationship of the set of primroses included in the set of flowers. Stedmon argues that understanding quantifiers in an extensional context requires not so much comprehension of class inclusion but rather comprehension of the extent of the domain of reference that is intended by the quantifier and that this interpretation of the scope of a quantifier can be shown to depend upon psychological, semantic and discourse factors. Children are undoubtedly easily misled in understanding the intended scope of quantifiers in extensional situations by various factors including those discussed in Chapter 2 - perceptual and conceptual salience. Presumably Stedmon regards intensional tasks as easier than extensional tasks because the intended scope of the quantifier is unambiguous in an intensional task: "All roses" would mean just that i.e. every single rose.

Stedmon's claim that the intensional asects of class inclusion are understood prior to the extensional aspects would seem to run counter to Piagetian theory which would claim that tasks which require an extensional understanding of the quantifier, i.e. concrete tasks, would be solved at an earlier age than intensional tasks which would seem to require an understanding of the logic of class inclusion rather than the empirical fact of one class included in another.

Markman (1978) makes a claim which runs counter to Stedmon's. She argues that although correct response on the standard kind of Piagetian class inclusion problem has frequently been taken as an index that the subject has mastered logical class inclusion, this interpretation of the subjects' responses is unwarranted since many subjects in their initial solutions to the problem base their responses on the empirical facts about the inclusion relation rather than an understanding of the logic of the inclusion relation. She argues in other words that the subjects can deal with the extensional aspect of inclusion prior to the intensional aspect. Markman would argue that for instance, in answering the question above: "Are there more flowers or more primroses?" the concrete operational subject would base his response on his empirical observation rather than deriving his response from consideration of the logical inclusion relation between flowers and primroses. Markman hypothesised that, in the absence of the empirical information on which to base their responses, subjects who could respond correctly on the standard class inclusion task (concrete operational subjects) would have difficulty in working out the answers to problems requiring some manipulation of the logical inclusion relation in the absence of empirical information on which to base their response: in one such task the subject had to decide whether the class inclusion relationship still holds if some of the items from the superordinate set are removed while in another task subjects were asked whether they could make it so that there were more of the subordinate elements (e.g. chairs) than the superordinate elements (e.g. furniture). Markman found that many concrete operational subjects had difficulty with this problem and other similar problems, and she interpreted this as consistent with her view that correct response on class inclusion problems does not entail that a subject understands the logic of class inclusion.

The good performance of even the youngest subjects in the class verification problem (80% correct overall) might be understood as an indication that subjects at this age can understand the logic of class inclusion and consequently that they should be able to perform well on a variety of tasks requiring an understanding of the logic of class inclusion. However as Markman's results have shown the ability to solve tasks involving class inclusion in the presence of an appropriate extensional situation does not guarantee the ability to solve class inclusion tasks in the absence of empirical support. Good performance on the verification of universally quantified statements in an extensional situation does not necessarily entail that the subject fully understands the logic of quantificational inclusion. However Stedmon on the other hand would presumably argue that the good performance on the verification task (an extensional task) would entail that subjects should have an understanding of the intensional aspects of inclusion.

Smith (1979) presented evidence that subjects from 4 to 7 years can respond correctly on tasks requiring comprehension of quantified inclusion and tasks involving an understanding of the logical consequences of this relation. Smith's tasks included:

(1) A quantified inclusion task: subjects were required to answer questions like:

(a) Are all boys people?
(b) Are all animals cats?
(c) Are all spoons bumblebees?

(2) Class and property inference tasks: subjects were required to answer questions such as:

(a) A pug is a kind of dog (animal). Does a pug have to be an animal (cat)? (Class inference)
(b) All flowers (roses) have stamens in them.
Do all plants (flowers) have stamens in them.

The items in the quantified inclusion task were like non-empirical (intensional) analogues of the universal affirmative propositions used in experiment 2. Rather than being verified empirically Smith's propositions were to be verified against knowledge in semantic memory. Questions like (a) corresponded to T type or subset relations, (b) to B type or superset relations and (c) to N type or disjoint relations respectively. E type or overlap relations were not included but would be something like "Are all women tories?". Stedmon would predict that young subjects should perform well on problems of this type since they involve understanding of an intensional aspect of class inclusion. Markman, on the other hand, would predict poor performance since response cannot be based on examination of an extensional array. Smith found good performance on the task even by the nursery group (mean age 4 years 8 months) with 82% correct. This high level of correct response provided some support for Stedmon's view that intensional tasks are easier than extensional tasks. However performance was very susceptible to presentation factors. Specifically performance was substantially worse for a group who had previously verified a block of questions involving the existential quantifier: in addition peformance was significantly worse on the second half of the blocks of questions. Errors seemed to occur either because subjects adopted the F.Q.P. type of interpretation and responded false to all three types of question since "all dogs are not all animals" or because subjects interpreted the quantifier as an existential quantifier and responded "Yes" to both subset and superset items and "No" on disjoint items. Smith argued that interpretation was initially determined by consideration of syntax but, as with the extensional tasks, was not stable and was liable to be influenced by pragmatic factors particularly after prolonged questioning. The kinds of errors made in verifying universally quantified propositions from semantic memory were similar to those made in verifiying universally quantified propositions against empirical arrays.

Smith also found better performance on the inference tasks than that usually found but again performance was highly susceptible to the influence of pragmatic factors.

VERIFICATION OF CONDITIONAL SENTENCES

RESULTS

A 6 (grade) X 4 (statement type) analysis of variance was carried out on correct response to conditional statements and the results of this analysis are shown in Table 3.7. Both factors, grade, $F(5,90)=15.31$, $p<0.001$, and statement type, $F(3,270)=66.60$, $p<0.001$, were significant and so was the grade X statement type interaction, $F(15,270)=6.94$, $p<0.001$.

Comparisons of simple interactions on adjacent grades showed that only the P6/S1 comparison was significant, $F(3,270)=4.44$, $p<0.01$ (Table 3.7a). Comparisons of simple interactions on non-adjacent grades showed that no primary grade comparisons were significant and only the S1/S5 secondary comparison, $F(3,270)=2.79$, $p<0.05$, was significant while all primary/secondary comparisons were significant.

The simple main effects analysis of statement type showed that statement type was significant at all primary grades and first year secondary but not at third or fifth year secondary (Table 3.7b).

Tests of simple main effects of grade showed significant F ratios only for E type statements, F(5,357)=1 . . , $p<0.001$, and B type statements, F(5,357)=1 . , $p<0.001$ (degrees of freedom calculated using

TABLE 3.7

ANALYSIS OF VARIANCE FOR CONDITIONAL STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
TOTAL	250.497	383			
BETWEEN	63.997	95			
GRADE	29.388	5	5.878	15.31	0.001
ERROR	34.609	90	0.384		
WITHIN	186.500	288			
STATEMENT TYPE	68.132	3	22.711	66.60	0.001
GS	32.915	15	2.194	6.94	0.001
ERROR	85.453	270	0.316		

TABLE 3.7a

TESTS OF SIMPLE INTERACTIONS FOR GRADE X STATEMENT TYPE INTERACTIONS FOR CONDITIONAL STATEMENTS: F VALUES AND SIGNIFICANCE LEVELS FOR BETWEEN GRADE COMPARISONS

	P2	P4	P6	S1	S3	S5
P2	-					
P4	0.165 ns	-				
P6	0.626 ns	0.461 ns	-			
S1	7.671 $p<0.001$	6.684 $p<0.001$	4.443 $p<0.01$	-		
S3	14.869 $p<0.001$	13.515 $p<0.001$	10.680 $p<0.001$	1.475 ns	-	
S5	14.701 $p<0.001$	12.460 $p<0.001$	11.503 $p<0.001$	2.794 $p<0.05$	0.823 ns	-

TABLE 3.7b

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF STATEMENT TYPE FOR CONDITIONAL STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
P2	33.563	3	11.188	35.40	0.001
P4	30.500	3	10.167	32.17	0.001
P6	26.625	3	8.875	28.09	0.001
S1	7.172	3	2.390	7.56	0.001
S3	1.875	3	0.625	1.90	ns
S5	1.313	3	0.438	1.39	ns
ERROR	85.453	270	0.316		

TABLE 3.7c

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF GRADE FOR CONDITIONAL STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
T	0.833	5	0.167	1.25	ns
N	0.083	5	0.017	0.29	ns
E	29.583	5	5.917	17.77	0.001
B	31.803	5	6.361	19.10	0.001
ERROR	120.002	360	0.333		

TABLE 3.7d

NEWMAN KEULS COMPARISONS ON STATEMENT TYPE MEANS FOR CONDITIONAL STATEMENTS

STATEMENT TYPE	B 0.948	E 1.083	T 1.667	N 1.979	r	s q(.95)(r,270) s q(.99)(r,270)
B 0.948	-	0.135 ns	0.719 **	1.031 **	4	0.160 0.211
E 1.083		-	0.584 **	0.896 **	3	0.191 0.239
T 1.667			-	0.312 **	2	0.210 0.257

the Satterthwaite approximation) (Table 3.7c).

Newman Keuls comparisons on means for statement types showed that all comparisons except that between B and E type statements were significant (see Table 3.7d).

As explained in the general introduction the sentence verification task was presented along with the picture verification, evaluation and syllogistic reasoning tasks in a within subjects design. The verification tasks were presented prior to the evaluation and syllogistic reasoning tasks to half of the subjects and for the other half the inference tasks preceded the verification tasks.

Since it was possible that experience with the inference tasks would lead to a facilitation of response on the verification tasks an analysis of order of presentation was carried out. Since levels of correct response on the class verification task were fairly high anyway it was thought that any improvement in performance when the verification task followed the inference tasks would be greater for conditional statements than for class statements. A 6 (grade) X 2 (linguistic form) X 2 (presentation order) analysis of variance was carried out on correct response. The results are shown in Table 3.8. Significant main effects were found for grade, $F(5,168)=74.57$, $p<0.001$, linguistic form, $F(1,168)=246.58$, p 0.001, and presentation order, $F(1,168)=11.32$, $p<0.001$. The grade X linguistic form interaction, $F(5,168)=6.30$, $p<0.001$, and the grade X presentation order interaction $F(5,168)=3.90$, $p<0.01$, were also significant.

The effect of presentation order reflected the fact that there were more correct responses when the verification task followed the inference tasks (82% correct) than when it was presented first (78%

TABLE 3.8

ANALYSIS OF VARIANCE FOR GRADE, LINGUISTIC FORM AND ORDER

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	359.479	191			
GRADE	154.354	5	30.871	74.57	0.001
LINGUISTIC FORM	102.083	1	102.083	246.58	0.001
ORDER	4.687	1	4.687	11.32	0.001
G X LF	13.042	5	2.608	6.30	0.001
G X O	8.063	5	1.613	3.90	0.01
LF X O	1.334	1	1.334	3.22	0.1
G X LF X O	6.291	5	1.258	3.04	0.1
ERROR	69.625	168	0.414		

correct). The grade x order interaction indicated that this was true only for subjects at certain grades, specifically only for subjects at primary 4, $F(1,168)=6.11$, $p<0.001$, and 5th. year secondary, $F(1,168)=9.03$, $p<0.001$. This unsystematic improvement in response is difficult to interpret. Neither the linguistic form X order nor the grade X linguistic form X presentation order interaction was significant indicating that presentation order effects were not greater for conditional statements.

CLASSIFICATION OF RESPONSE PATTERNS

Table 3.6, p.161, shows the number of subjects at each grade who responded according to the different interpretations of the conditional - conjunction, biconditional and conditional. A modified classification of response patterns was also carried out but results on this will only be reported where they differ from the standard classification.

The predominant consistent response pattern of primary children was the conjunctive response: about half the primary school children responded according to this pattern. No primary children responded according to the conditional interpretation. The advent of a conditional interpretation at first year secondary was accompanied by a decrease in conjunctive and an increase in inconsistent response.

No adjacent grade comparisons of response pattern distributions were significant although all primary/secondary comparisons except the P6/S1 were significant on a chi-square test as was the S1/S5 comparison, $\chi^2(3)= 6.984$, $p<0.05$.

DISCUSSION

The simple main effects analysis of grade for statement types confirmed what an inspection of Figure 3.2b suggests: improvement in performance with increasing age was due only to improvement on E and B type statements. Performance on N and T type statements did not change significantly across grade. Primary school subjects performed very poorly on B and E type statements: only at first year secondary did subjects begin to realize that these statements were false. The significance of the test of simple interactions on the P6/S1 comparison reflected the increase in correct response only on B and E type statements, but the response pattern classification showed that there was only a gradual change in response pattern with still only two subjects at first year secondary classified as conditional. Considering the two analyses together it seems that there is a general increase in the realization that B and E type statements are false at first year secondary but subjects are still at a transitional stage and their responses are still rather inconsistent. Interestingly, despite the transitional nature of response at this stage, very few biconditional response patterns were found. The further non-significant increase in correct response on B and E type statements from S1 to S3 was accompanied by an increase in the predominance of the conditional response pattern indicating that performance on the task was stabilising. The absence of a simple main effect of statement type for S3 and S5 subjects indicated that by S3 performance on the conditional verification task was mature.

The results of the conditional verification task replicated those of Kuhn's conditional reasoning task showing the predominance of the conjunctive interpretation by primary school children and the change to conditional interpretation around 12 years. The age at which the conditional interpretation appears is consistent with Kuhn's view that the conditional verification task is a formal operational task involving comprehension of implication.

The significance of the simple effects of linguistic form at all grades up to first year secondary reflected the superior performance on class verification compared with conditional verification at these grades while the significance of the simple interaction of linguistic form and statement type at these grades suggested that the superior performance on class statements was specific to statement types B and E. The superior performance on class statements and the significant increase in correct response on conditional verification from P6 to S1 are consistent with Piaget's claim that concrete operational subjects should be able to deal with arguments stated in class and relational logic but not until formal operations should they be able to deal with propositional logic arguments.

It could be argued that a result showing no difference in performance on class and conditional statements would also have been compatible with Piaget's theory since, as mentioned in the introduction to this chapter, Piaget's theory makes ambiguous predictions about the abilities of concrete operational subjects to comprehend propositional relationships. Inhelder and Piaget argued that the language of a subject at a particular operational level is not a reliable guide to his underlying cognitive operations. Since a subject's linguistic comprehension is also constrained by the operational structures available rather than by the linguistic expression of the logical relationship no difference in performance might be predicted between the logically equivalent class and conditional statements.

However the results of experiment 2 indicated that for subjects at the concrete operational stage the linguistic expression of the logical relationship is an important factor in determining how the logical relationship is understood. When universal quantification is expressed linguistically in a form which corresponds more naturally to class inclusion, i.e. as "All A are B", the younger subjects can understand it, whereas when universal quantification is expressed linguistically using the general conditional, "If x is an A then x is a B" the younger subjects have difficulty with it. Presumably this difference in performance with the different linguistic expressions would be explained within the Piagetian framework in terms of the subject mapping the class statement onto the classification operations of concrete operations but failing to map the conditional onto the classification operations and not yet having the appropriate propositional operations to map the conditional onto. Only once the subject has acquired the formal operational structures would the mapping from the conditional to the appropriate cognitive structures be possible.

Although the conditional reasoning task has been discussed as though it involved a propositional conditional in fact it involved a general conditional. The general conditional involves quantification which the propositional conditional does not and it is consequently more complex than the propositional conditional. It is possible that the younger children give incorrect responses on the conditional verification task not because of a failure to understand implication resulting from the lack of formal operational structures, but because the young subjects do not understand the specific use of the conditional in conjunction with the indefinite article "a/an" as corresponding to universal quantification (Quine, 1941/80). Certainly a subject who did understand the conditional as corresponding to the universal quantifier should be able to respond on the conditional statements in exactly the same way as on the class statements. Subjects who did not understand the device of the conditional plus the indefinite article as corresponding to a universal quantifier are quite likely to understand the general conditional as a statement about a specific entity. A statement like "In my garden if an insect is big it is black.", for example, will probably be understood as an existential statement about a big black insect, i.e. "In my garden there is big black insect.". It is possible that some of the difficulties of the younger subjects on the conditional verification task are due to the kinds of difficulty described by Stedmon (1982) in establishing the relevant reference sets for quantifiers. The use of the indefinite article, a/an, in the general conditional probably makes it particularly difficult to identify the appropriate reference sets in the extensional task. However difficulty in establishing the relevant reference sets would seem to stem from a more fundamental difficulty in understanding the linguistic convention. According to the linguistic explanation the significant increase in correct response on the conditional verification task at from primary to secondary would simply be explained in terms of the acquisition of the ability to understand "conditional plus indefinite article = universal quantifier" rather than entailing a re-organisation of cognitive structure.

Although failure to comprehend the quantificational aspect of the general conditional may be what causes the problems in the conditional verification task, consideration of how an analogue of the conditional verification task involving a propositional conditional rather than a general conditional might be constructed reveals that it would be difficult to construct such a task since, as was mentioned in the introduction, a propositional conditional is a statement of a specific relationship and is true when any one of three different conditions holds (pq, -pq or -p-q) whereas a general conditional is a general rule. For example the general conditional "In my garden if an insect is big it is striped." is a rule which applies to every insect in the specified domain (in my garden) and states that for every insect in the domain the relationship between its size and marking can be described in terms of a conditional relationship. A propositional conditional similar to this general conditional would be e.g. "In my garden if the insect(s) is(are) big then it is(they are) striped." A verification task like that above with a propositional conditional would not make sense since the conditional cannot be evaluated as true or false with respect to three different types of insect simultaneously. The evaluation of the truth of a propositional conditional with respect to specific exemplars is discussed in Chapter 5. Since is not possible to carry out an analogue of the verification task with a propositional conditional it is not possible to say whether the difficulty of the younger subjects on the verification task with the general conditional is due to their inability to comprehend the implication relation or their lack of comprehension of the general conditional construction as corresponding to universal quantification. There is however evidence from other studies showing that young children have difficulty in comprehending implication.

The results of the class and conditional verification problems in experiment 2 are, as we have seen, consistent with the broad framework of Piaget's theory. The different response patterns of primary school children on the verification of class and conditional statements clearly cause problems for propositional models of verification, like the comparison model, which would propose that errors in verification are a function of propositional structure. Since the universally quantified statement "All A are B" and the general conditional "If x is an A then x is a B." have the same propositional representations : $\forall x(Ax \rightarrow Bx)$, a propositionally based theory would predict the same patterns of error on these statements. The analysis of simple interactions for the statement type by linguistic form interaction in the overall analysis of variance showed that only at third and fifth year secondary was the interaction non-significant. The predictions of a propositionally based theory are not upheld for subjects younger than third year secondary.

An interesting divergence of opinion emerges in the literature on information processing models of language comprehension and deductive inference on the one hand and the development of reasoning abilities on the other, concerning the relative difficulties of sentence verification tasks and tasks involving deductive reasoning. The similarities between the operations proposed in recent models of deductive reasoning (Clark, 1969; Revlis, 1975; Mayer and Revlin, 1978 etc.) and recent models of sentence verification (Chase and Clark, 1972; Trabasso et al, 1971 etc.) has been emphasised by Revlis (1975) and Evans (1982). Figures 3.6a and 3.6b illustrate the similarities in the sequences of operations in the models. Both models propose that the information from the two separate sources - either premises and conclusion or sentence and picture are encoded in similar formats. These representations are then compared in a systematic way.

The models differ in that the reasoning model includes a stage at which a composite representation of the information from the premises is formed. Since task difficulty is predicted to be a function of the complexity of the model it would be reasonable to predict that tasks involving deductive inference would be more difficult than sentence verification tasks. It will be recognised however that this is

Figure 3.6

Comparison of Sentence Verification (a) and

Deductive Reasoning (b) models

	(a)	(b)
Stage 1	Encode sentence	Encode premises
	↓	↓
Stage 2	Encode picture	Encode conclusion Form composite representation
	↓	↓
Stage 3	Comparison process	Comparison process
	↓	↓
Stage 4	Decision/response	Decision/Response

contrary to Kuhn's proposals concerning the relative difficulty of verification and syllogistic reasoning. She claimed that syllogistic reasoning ability is a concrete operational ability and is consequently acquired prior to the ability to verify conditional statements which is a formal operational acquisition. Although the results of experiment 2 are in accordance with Kuhn's claims there are problems with her claim that syllogistic reasoning is a concrete operational ability. These will be discussed in Chapter 6.

PICTURE VERIFICATION

In the sentence verification task the subject was presented with a picture depicting an inclusion/implication relationship and was required to say whether eight different class or general conditional sentences were true or false with respect to the picture. Although primary school subjects performed well with the class statements they performed very poorly with certain types of conditional sentence since they seemed to interpret the conditional as a conjunction. Models of sentence verification like the comparison model propose that when the sentence is presented prior to the picture, the encoding of the picture depends upon the terms in the sentence (Revlin and Leirer, 1980). In encoding the picture the relationship represented in the picture between the terms of the sentence must be determined. In experiment 2 the picture was presented prior to the sentences to be verified. It is possible that, since younger subjects tend to focus on perceptual information, they tended to encode the information from the picture in a fixed way and did not reconsider the representation of the picture in the light of each sentence. It was thought that changing the picture rather than the sentences might encourage the younger subjects to consider each sentence/picture pair more carefully and improve response, particularly on conditional verification.

EXPERIMENT 3

THE PICTURE VERIFICATION TASK

METHOD

The task was similar to that in experiment 2 in that the subjects were required to say whether certain sentence/picture combinations were true or false. The difference was that whereas in experiment 2 the subject had to say whether the eight different implication/inclusion sentences which can be made by combining positive and negative values of two binary propositions p and q were true or false with respect to a particular picture depicting an implication relationship, in experiment 3 the subject had to say whether a particular implication/inclusion sentence was true or false with respect to the four different "implication" pictures which can be made from combinations of three different kinds of element. For instance, the pictures to be verified against a sentence such as "In my garden if an insect is big it is striped." are shown in Figure 3.7. As in experiment 2 the instructions for the problem were contained within a story. The story for the insect problem was as follows (that for shape content is in Appendix B):

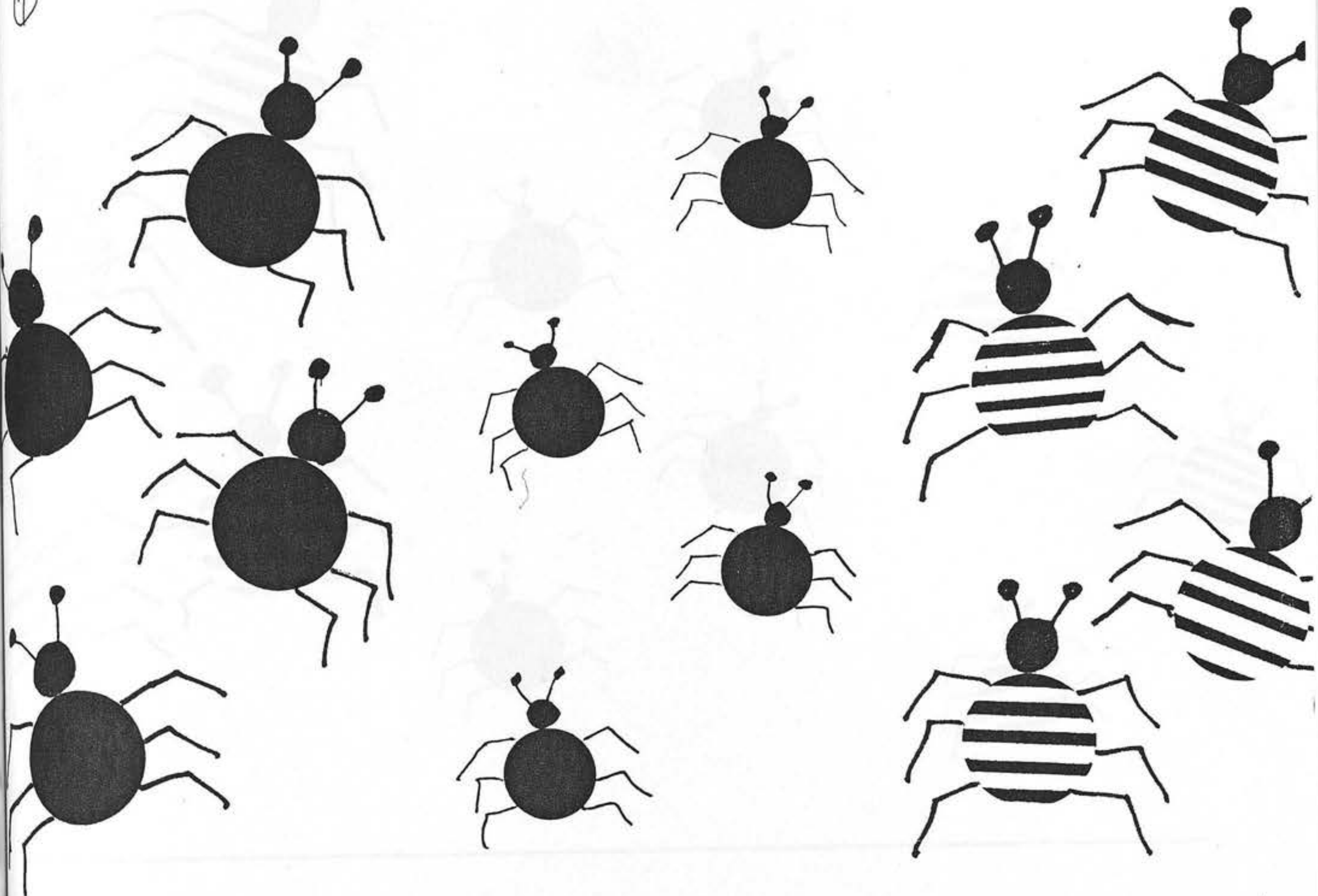
"One day when Mr. Jones was in his garden he noticed lots of insects. He watched carefully and saw three different kinds of insect. Later he was telling a friend about the insects in his garden and he said "In my garden if an insect is big it is striped". Is this sentence true for these insects....these insects? etc.".

The subject was then presented with the four different implication pictures one at a time in random order and had to say whether the

Figure 3.7

"Implication" Pictures for picture verification task

1



2

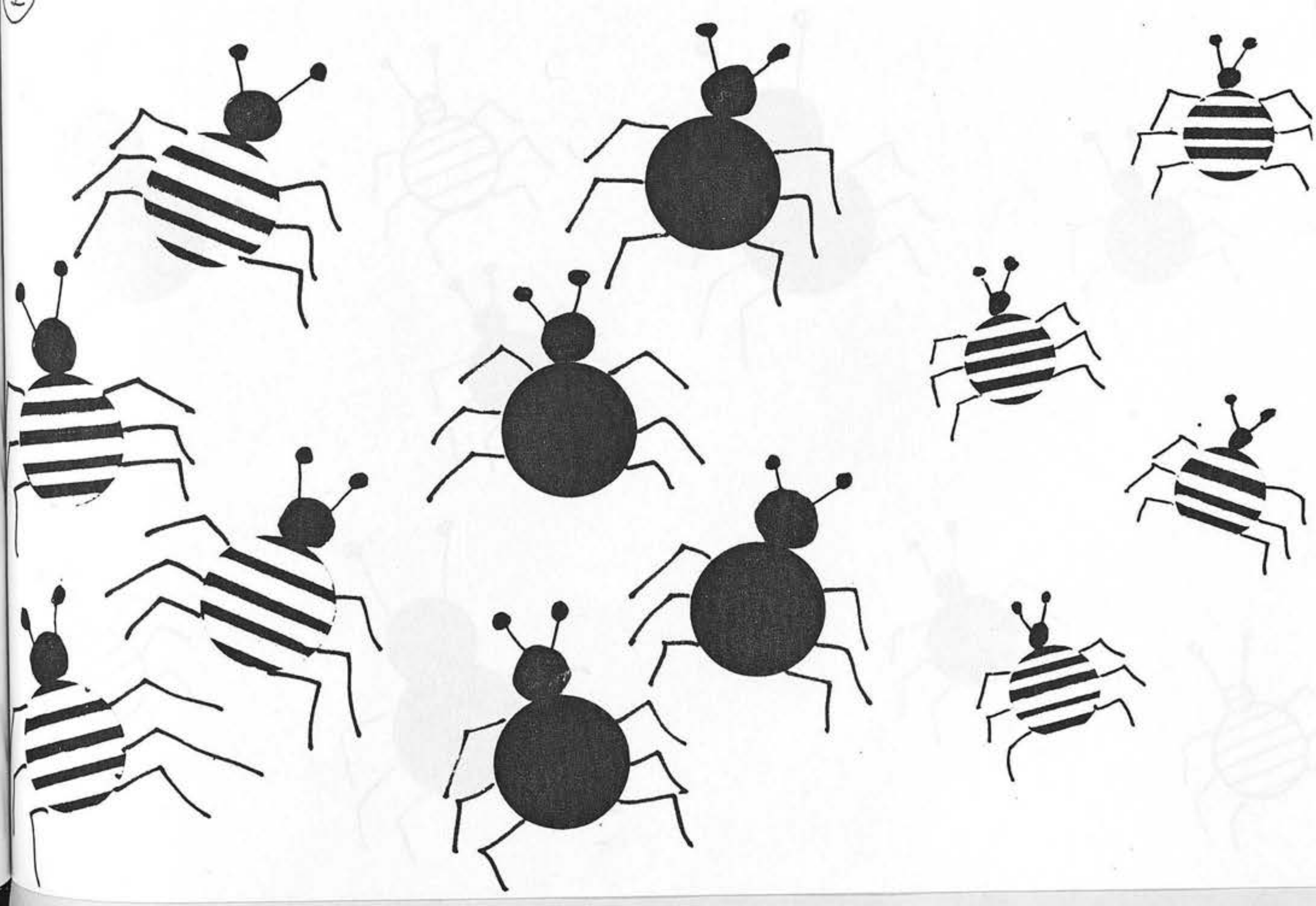
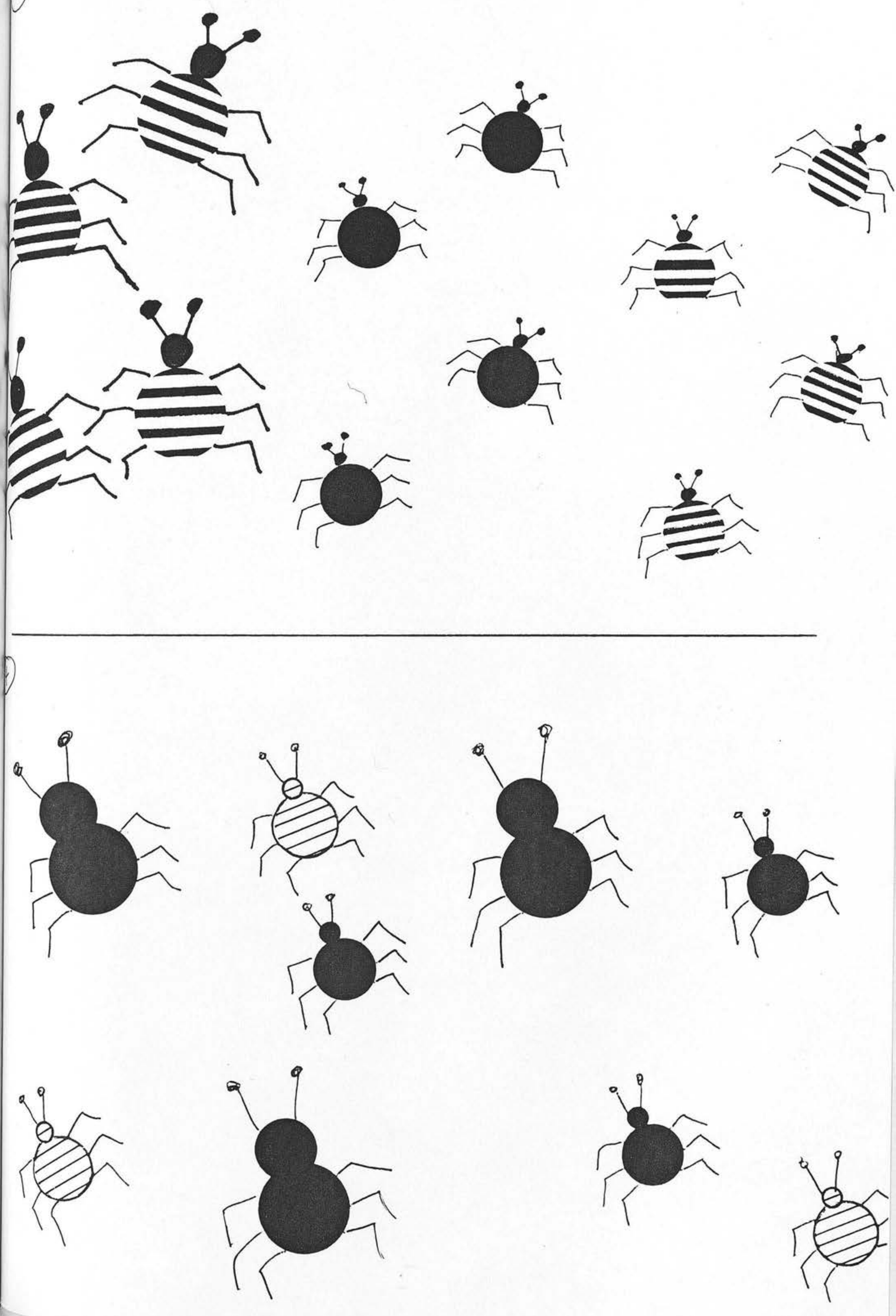


Figure 3.7 (cont.)



sentence was true for each picture in turn. Half of the subjects were required to verify class statements and the other half conditional statements. Subjects dealt with statements of the same linguistic form in experiments 2 and 3 but those subjects who received insect content in experiment 2 received shape content in experiment 3 and vice versa.

RESULTS

Table 3.9 shows percent correct response as a function of grade, linguistic form of statement and statement type while Table 3.10 shows the results of an analysis of variance carried out on correct response. The results were similar to those observed in the sentence verification task except that the grade X linguistic form interaction did not reach significance since correct response on class statements was superior to that on conditional statements at all grades. Overall levels of correct response on experiments 2 and 3 were similar: Overall 87% of responses to class statements on experiment 3 and 88% on experiment 2 were correct cmpared with 70% of responses to conditional statements on experiment 3 and 71% on experiment 2. Following Sternberg's (1979, p. 485) method for comparing the difficulty of two tasks a paired t-test was carried out on the mean proportion of correct responses collapsed across linguistic form and statement type on the sentence and picture verification tasks for subjects at each grade. The t-test for related measures, $t(5) = 1.34$, indicated that mean correct responses on sentence and picture verification tasks were not significantly different.

Table 3.1l shows the distribution of response patterns across grade for the picture verification task. Broadly speaking the results were similar to those of the sentence verification task although overall

TABLE 3.9

MEAN PERCENT CORRECT RESPONSE ON DIFFERENT STATEMENT TYPES (T, B, E, N) FOR DIFFERENT LINGUISTIC FORMS (CLASS AND CONDITIONAL) ACROSS GRADE FOR THE PICTURE VERIFICATION TASK

	T			B			E			N			TOT		
	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT
P2	69	81	75	56	38	47	69	19	44	100	94	97	73	58	66
P4	69	100	84	56	6	31	75	6	41	100	100	100	75	53	64
P6	81	81	81	88	25	56	81	31	56	100	100	100	88	59	73
S1	88	75	81	94	69	81	88	69	78	100	100	100	92	78	85
S3	81	88	84	94	94	94	94	69	81	100	100	100	92	88	90
S5	100	75	82	100	81	91	94	88	91	100	100	100	98	86	92
TOT	81	83	82	81	52	67	83	47	65	100	99	99	87	70	78

TABLE 3.10

ANALYSIS OF VARIANCE FOR GRADE, STATEMENT TYPE AND LINGUISTIC FORM FOR PICTURE VERIFICATION TASK

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	130.120	767			
GRADE	43.120	5	1.946	12.80	0.001
LINGUISTIC FORM	5.005	1	5.005	32.93	0.001
LF X G	1.036	5	0.207	1.36	ns
ERROR	27.35	180	0.152		
WITHIN	87.000	576			
STATEMENT TYPE	14.860	3	4.953	50.03	0.001
ST X G	8.156	15	0.544	5.49	0.001
ST X LF	5.484	3	1.828	18.46	0.001
ST X LF X G	4.850	15	0.323	3.26	0.001
ERROR	53.65	540	0.099		

TABLE 3.11

CLASSIFICATION OF RESPONSE PATTERNS FOR PICTURE VERIFICATION TASK

CLASS

	CONJ	BICOND	COND	INCONSISTENT
P2	3	1	4	8
P4	2	4	6	4
P6	1	1	10	4
S1	0	1	12	3
S3	1	0	12	3
S5	0	0	15	1
TOTAL	7	7	59	23

CONDITIONAL

	CONJ	BICOND	COND	INCONSISTENT
P2	6	1	2	7
P4	15	0	1	0
P6	10	0	3	3
S1	3	0	8	5
S3	0	0	9	6
S5	2	0	10	4
TOTAL	36	1	33	26

TOTAL

	CONJ	BICOND	COND	INCONSISTENT
P2	9	2	6	15
P4	17	4	7	4
P6	11	1	13	7
S1	3	1	20	8
S3	1	0	21	10
S5	2	0	25	5
TOTAL	43	8	92	49

more response patterns were classified according to a consistent response pattern in the picture verification task (74%) than in the sentence verification task (55%). A t-test on related measures was carried out on mean proportion of correct conditional/inclusion response patterns to the sentence and picture verification tasks and the significance of this test, $t(5)=3.93$, $p<0.05$, reflected the higher number of correct response patterns on the picture verification task (48%) compared with the sentence verification task (37%). However this probably reflected nothing more interesting than the fact that the picture verification task included only four items while the sentence verification task included eight items and consequently subjects were more likely to be classified consistent on the picture verification than on the sentence verification task.

Despite the difference in the incidence of consistent response patterns between experiments 2 and 3, the predominant response patterns for subjects at different grades in experiment 3 were similar to those in experiment 2. The predominant response pattern in the class verification task at all grades was the class inclusion/implication pattern: 61% of patterns overall were of this type compared with 54% in experiment 2. There was a gradual increase in correct response with increasing age. As in experiment 2 the predominant consistent response pattern in the conditional verification task for primary subjects on experiment 3 was the conjunctive response but this changed to the conditional for secondary subjects. Overall there were more conditional response patterns on experiment 3 (34%) than on experiment 2 (9%).

The analysis of mean correct response to the sentence and picture verification tasks showed that changing the pictures to be verified rather than the sentences in a sentence picture verification task did not affect the levels of correct response. Since the sentence verification task has already been discussed the picture verification task is not discussed further.

SUMMARY

Experiment 2 extended the class verification task to include all possible class inclusion relations which can be made by combining positive and negative values of two binary attributes p and q. Performance on this class verification task by subjects from 6 to 17 years was compared with that on a similar task described by Kuhn (1977) involving the verification of conditional statements. Since the class and general conditional statements are logically equivalent, propositional models of verification which are based on the logical form of the statements would predict no difference in performance on verifying class and conditional statements. 88% of responses on the verification of class statements were correct and even the youngest subjects performed well. As in experiment 1 the errors they did make were not compatible with Bucci's "structure neutral" hypothesis but seemed to be related instead to aspects of the presentation of the task. The results of the conditional verification problem replicated Kuhn's results showing that children under 12 years have specific difficulties with the conditional verification problem. The pattern of errors on different statement types indicated that children of this age seemed to interpret the conditional as a conjunction. Significant increases in correct response were found for secondary subjects. These results are compatible with a simple interpretation of Piaget's theory which would hold that the concrete operational subject can deal with linguistic statements such as the class inclusion statement which map in a relatively intuitive way onto the underlying operations of class logic but not until formal operations can he deal with propositional logic statements.

Experiment 3 involved a procedural variation of the verification task

in which subjects were required to say whether a particular class inclusion/ implication statement was true or false with respect to four different empirical instantiations of an inclusion/implication relationship. Levels of correct response on the different statement types were essentially similar to those found in experiment 2.

CHAPTER 4

VERIFICATION WITH DIFFERENT LINGUISTIC VARIANTS OF THE UNIVERSAL QUANTIFIER.

In experiments 2 and 3 subjects up to and including first year secondary (12 years) were found to perform better on a verification task with class statements of the type "All A are B" than with the logically equivalent general conditional statements of the type "if p then q". The superior performance on class statements indicated that propositionally based models of sentence verification which propose that errors in verification are a function of propositional representation were inadequate as explanations of the errors of primary school subjects since the propositional representations of class and conditional statements are identical. The superior performance on class statements was mainly due to superior performance on E and B type statements. The results of the statement type analysis, along with the response pattern data, indicated that while primary school subjects interpreted the class statements correctly they tended to interpret the general conditional as a conjunction.

Kuhn argued that the difficulty the younger subjects had in verifying the conditional reflected the fact that they had not yet acquired formal operational thinking which provides the operations necessary for interpreting empirical data in terms of propositional relationships. It was suggested that an alternative explanation of the younger subjects' difficulties with the conditional was that the younger subjects did not understand the linguistic device of the conditional plus the indefinite article as corresponding to universal quantification. In experiments 2 and 3 class and conditional statements were presented in a between subjects design. It was considered possible that, if the difficulty of the younger subjects in interpreting the general conditional was due to their inability to understand the general conditional as corresponding to the universal quantification, a within subjects presentation of class and conditional statements would bring out the similarity in meaning for the young subjects. On the other hand if the difficulty of the younger subjects in verifying the general conditional reflected the absence of the relevant cognitive structures a within subjects design would not be expected to improve performance on conditional verification. In experiment 4 class and conditional statements were presented in a within subjects design in order to see whether this led to an improvement in performance of the younger subjects in verifying E and B type conditional statements.

There are many natural language translations of universal quantification - "all", "every", "each", "any" "everything" the general conditional and even "a/an" (Quine, 1941/80) although these different natural language variants have differences in meaning which the quantificational form alone does not capture (Cresswell, 1973; Fodor, 1982). Although "any" is considered as a natural language translation of universal quantification it also has much in common with the existential quantifier "some". Unlike "all", for example, which takes a plural verb "any" takes a singular verb. It is possible that young children would be liable to misinterpret "any" in the same kind of way as they misinterpret the general conditional. The quantifier "every" is like "any" in that it takes a singular verb but, although apparently similar in meaning "any" and "all" have different scope properties in certain circumstances or different "acceptablity principles" (Cresswell, 1973). It was thought that "every" might be intermediate in difficulty between "any" and "all" since it seems to have properties of both. It was decided to look at how the quantifiers "any" and "every" are verified as well as "all" and "if then". In experiment 4 subjects were required to verify universally quantified statements expressed with the linguistic quantifiers "all", "every", "any" and the general conditional "if-then".

In a study of the development of understanding of the logical connectives using a type of verification task Neimark and Slotnick (1970) found that younger subjects (around 8 years) produce more correct responses when pictorial elements were to be "verified" while older subjects (up to 13 years) produced more correct responses on verbal items. In experiment 4 half the subjects were presented with a pictorial representation of the elements as in the previous verification studies while the other half were presented with a verbal description of the elements e.g. red squares, blue squares, red circles.

EXPERIMENT 4

VERIFICATION WITH VARIOUS UNIVERSALLY QUANTIFIED EXPRESSIONS

METHOD

Subjects

32 subjects from primary 1, primary 3 and primary 5 of an Edinburgh primary school took part. Mean ages (and age ranges) were 6 years 0 months (5 years 6 months to 6 years 3 months), 7 years 11 months (7 years 4 months to 8 years 3 months) and 9 years 11 months (9 years 5 months to 10 yrs. 7 months).

Procedure

The task was essentially similar to the sentence verification task of experiment 2. In experiment 4 however the subject was required to verify 32 statements altogether comprising the eight possible quantified inclusion or implication relationships for four different linguistic expressions of the implication relationship: these were "if then" "every" "all" and "any".

Two different content areas were used - shapes and insects. Half the subjects verified statements with respect to pictures depicting an inclusion/implication relationship while the other half verified statements with respect to verbal descriptions of the pictures e.g."RED SQUARES, BLUE SQUARES, RED CIRCLES". Content and mode (verbal/pictorial) were combined factorially so that eight subjects at each grade received a particular content and mode.

Presentation of the different statement types was blocked by quantifier and a different picture (or verbal description) was used with every quantifier for any particular subject. For any particular quantifier the four different verbal/pictorial representations were presented four times. The order of presentation of quantifier was random as was the order of presentation of the different statement types, T, B, E and N.

In order to provide a context for the problem a short story was read. For shape content the problem was as follows (The story for insect content is in Appendix C):

"Fred's teacher was giving his class a lesson on shapes. The teacher gave every child a card with coloured shapes on it. This is the card that Fred had. On Fred's card there were....(DESCRIPTION OF SHAPES ON CARD)......Later on the teacher was trying to find out how much the children remembered about the lesson. She picked up Fred's card and asked the children to say which of the following sentences were true about the shapes on Fred's card. Can you do it?"

The child was presented with a pictorial or verbal representation of the shapes. His task was to say for each sentence whether it was true or false with respect to the verbal/pictorial representation of shapes.

RESULTS

Table 4.1 shows percent correct response to the 4 different statement types for the four different linguistic forms for subjects at different grades. A 3 (grade) X 4 (linguistic form) X 4 (statement type) analysis of variance with repeated measures on linguistic form

TABLE 4.1

MEAN PERCENT CORRECT RESPONSE AS A FUNCTION OF GRADE, QUANTIFIER AND STATEMENT TYPE

IF THEN

	P1	P3	P5	TOT
T	73	73	73	73
B	29	48	56	44
E	48	48	69	60
N	81	100	98	93
TOT	58	71	74	68

ALL

	P1	P3	P5	TOT
T	73	75	71	73
B	44	52	67	54
E	65	77	85	76
N	88	98	98	94
TOT	67	75	80	74

EVERY

	P1	P3	P5	TOT
T	60	77	71	69
B	48	63	65	58
E	77	77	83	79
N	90	100	94	94
TOT	69	79	78	75

ANY

	P1	P3	P5	TOT
T	75	71	75	74
B	35	50	65	50
E	49	52	71	56
N	81	96	94	90
TOT	59	67	76	67

TOT

	P1	P3	P5	TOT
T	70	74	72	72
B	39	53	63	52
E	58	67	77	68
N	85	98	96	93
TOT	63	73	77	71

and statement type was carried out on correct response. The results of this analysis are shown in Table 4.2. The significant main effect of grade, $F(2,69)=8.77$, $p<0.001$, reflected the general increase in correct response with increasing age. A significant main effect was also found for type of quantifier, $F(3,207)=7.02$, $p<0.001$, reflecting different levels of correct response with different quantifiers. The significant main effect of statement type $F(3,207)=46.00$, $p<0.001$, showed that performance on some statement types was better than that on others: specifically performance on N (93% correct) was better than that on T (72% correct) which was better than that on E (68% correct) which was better than that on B (52% correct). The significant quantifier X statement type interaction, $F(9,621)=3.22$, $p<0.005$ indicated that this level of correct response on the different statement types did not hold for all four quantifiers.

Comparisons between means for grade were calculated using Tukey's HSD test (Table 4.2a). The tests showed that primary 3 subjects (73% correct overall) scored significantly better than primary 1 subjects (63% correct overall), but the difference between primary 3 and primary 5 subjects (77% correct overall) was not significant (Table 4.2a). Since grade did not interact significantly with statement type or quantifier it can be presumed that the effects of statement type and quantifier held across grade.

The significant main effect of type of quantifier indicated that subjects performed better with some quantifiers than others: specifically subjects had more correct responses with "all" (74% correct overall) and "every" (75% correct overall) than they did with "if-then" (68% correct overall) and "any" (67% correct overall). Tukey's HSD was calculated for comparisons between means for

TABLE 4.2

ANALYSIS OF VARIANCE FOR QUANTIFIER, GRADE AND STATEMENT TYPE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	611.124	1151			
BETWEEN	78.312	71			
GRADE	15.866	2	7.933	8.77	0.001
ERROR	62.446	69	0.905		
WITHIN	532.812	1080			
QUANTIFIER	6.336	3	2.112	7.02	0.001
Q X G	1.307	6	0.218	0.72	0.001
ERROR 2	62.419	207	0.301		
STATEMENT TYPE	100.336	3	33.445	46.00	0.001
S X G	6.245	6	1.041	1.43	ns
ERROR 3	150.481	207	0.727		
S X Q	9.049	9	1.005	3.22	0.005
S X Q X G	2.777	18	0.154	0.49	ns
ERROR 4	193.862	621	0.312		

TABLE 4.2a

DIFFERENCES BETWEEN MEANS FOR BETWEEN GRADE COMPARISONS

GRADE	P1 1.263	P3 1.463	P5 1.542
P1 1.263	-	0.200 *	0.279 **
P3 1.463		-	0.079 ns

ERROR TERM FOR TUKEY'S RATIO FOR COMPARISONS BETWEEN GRADES =
(MS(subjects within groups)/nqr) = (.905/(24 x 4 x 4)) = 0.048

Tukey's HSD = q(0.05)(3,69) X ERROR = 3.69 X 0.048 = 0.163

q(0.01)(3,69) X ERROR = 4.27 X 0.048 = 0.205

quantifiers collapsed across statement type (Table 4.2b). Comparisons between all quantifiers except between 'if then' and 'any' and between 'all' and 'every' were significant.

The significant quantifier X statement type interaction however showed that the difference between quantifiers was only true for particular statement types. Tests of simple main effects of quantifier for the different statement types showed that significantly different levels of correct response across quantifier were found only for statement type B, $F(3,736)=3.27$, $p<0.05$ and statement type E, $F(7,736)=12.59, p<0.001$ (Table 4.2c).

Following Kirk (1968) post hoc comparisons between quantifiers for statement types B and E were tested against the pooled error term derived from the F ratio denominator used for testing MS(B) at c(i) (Kirk, 1968, p.306) (see Table 4.2d). Following Kirk (p.269) an approximation of the critical value of q' for calculating Tukey's HSD when pooled error terms are involved was calculated (see Table 4.2d). Tukey's HSD test on comparisons between quantifiers for statement type B showed that only the comparison between 'if then' (44% correct) and 'every' (58% correct) was significant at the 0.05 level. Post hoc comparisons between quantifier for statement type E showed significant differences between all comparisons except between 'if then' and 'any' and between 'every' and 'all'.

Analysis of simple main effects of statement type showed that the effect of statement type was significant for all quantifiers (Table 4.2e). Tukey's HSD tests on comparisons between statement types were carried out for each quantifier using the pooled error term derived from the F ratio denominator for testing MS(C) at b(i) (see Table 4.2f). An approximation of the critical value of q' for pooled error

TABLE 4.2b

COMPARISONS BETWEEN MEANS FOR QUANTIFIERs

	ANY 1.347	IF THEN 1.351	ALL 1.486	EVERY 1.507
ANY 1.347	-	0.004 ns	0.139 *	0.160 **
IF THEN 1.351		-	0.135 *	0.156 **
ALL 1.486			-	0.021 ns

ERROR TERM FOR TUKEY'S RATIO FOR COMPARISONS BETWEEN QUANTIFIERS
$=(\text{MS(B X subjects within groups)}/npr)^{\frac{1}{2}} = (0.301/24 \times 3 \times 4)^{\frac{1}{2}} = 0.032$

Tukey's HSD = q(0.05)(4,207) X ERROR = 3.65 X 0.032 = 0.117

q(0.01)(4,207) X ERROR = 4.27 X 0.032 = 0.142

TABLE 4.2C

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF QUANTIFIER AT STATEMENT TYPE (i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
T	0.305	3	0.102	0.33	ns
B	3.042	3	1.014	3.27	0.05
E	11.705	3	3.902	12.59	0.001
N	0.333	3	0.111	0.36	ns
ERROR	256.281	828	0.309		

TABLE 4.2d

COMPARISONS BETWEEN MEANS FOR QUANTIFIERS AT STATEMENT TYPE(i)

An approximation of the critical value of q for the comparison between quantifiers at statement type (i) is calculated by:

q(0.05) = q(4,207)MS(B x subj.within groups) + q(4,621)MS(BC x subj. within groups)(r-1)/MS(B x subj. within groups) + MS(BC x subj. within groups) (r-1) =3.67(0.301) +3.66(.312)(4-1)/0.301 + (0.312)(4-1)= 3.66
q(0.01) = 4.46(0.301) + 4.44(0.312)(4-1)/0.301 + (0.312)(4-1) = 4.44

Error term for calculating Tukey's HSD for comparisons between quantifiers: = (MS(B x subjects within groups) + MS(BC x subjects within groups)(r-1)/rpn)$^{1/2}$ = (.309/24 x 3)$^{1/2}$ = 0.065
Tukey's HSD = q(0.05) x ERROR = 3.66 X 0.065 = 0.238
q(0.01) x ERROR = 4.44 X 0.065 = 0.289

COMPARISONS BETWEEN MEANS FOR QUANTIFIER FOR E TYPE STATEMENTS

	ANY 1.111	IF THEN 1.194	ALL 1.514	EVERY 1.583
ANY 1.111	-	0.083 ns	0.403 **	0.472 **
IF THEN 1.194		-	0.320 **	0.389 **
ALL 1.514			-	0.069 ns

COMPARISONS BETWEEN MEANS FOR QUANTIFIER FOR B TYPE STATEMENTS

	IF THEN 0.889	ANY 1.000	ALL 1.083	EVERY 1.167
IF THEN 0.889	-	0.111 ns	0.194 ns	0.278 *
ANY 1.000		-	0.083 ns	0.167 ns
ALL 1.083			-	0.084 ns

TABLE 4.2e

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF STATEMENT TYPE AT QUANTIFIER(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
IF THEN	36.705	3	12.235	29.41	0.001
ALL	23.472	3	7.824	18.81	0.001
EVERY	20.264	3	6.755	12.59	0.001
ANY	28.944	3	9.648	0.36	0.001
ERROR	344.343	828	0.416		

terms was calculated (see Table 4.2f).

For 'if then', comparisons between all statement types except T and E were significant and order of difficulty was N T E B.

For 'all', comparisons between all statement types except T and E were significant. The order of difficulty was N E T B.

For 'every', comparisons between all statement types except between T and B and between T and E were significant and order of difficulty was as for 'all', N E T B.

For 'any', all comparisons except the B/E comparison were significant. Order of difficulty was as for 'if then', N T E B.

In order to establish whether either content or presentation mode influenced response two further analyses were carried out. To simplify analysis the data were collapsed across statement type and a 3 (grade) X 2 (mode) X 4 (quantifier) analysis of variance with repeated measures on quantifier type was carried out on correct response (Table 4.3). The results show the expected significant effects of grade and quantifier type. However neither the effect of presentation mode nor any of the interactions of this factor were significant.

A 3 (grade) X 2 (content) X 4 (quantifier) analysis of variance was carried out on the data (Table 4.4). Significant effects were found for grade and quantifier type. The main effect of content was not significant but the grade X quantifier type X content interaction just reached significance at the 0.05 level. This seemed largely attributable to superior performance by P3 subjects on shape content and superior performance on insect content by P5 subjects for the quantifier "any" but little difference in performance on the two

TABLE 4.2f

COMPARISONS BETWEEN MEANS FOR STATEMENT TYPES AT QUANTIFIER (i)

An approximation of the critical value of q for the comparison between different statement types at quantifier (i) is calculated by:

q(0.05) = q(4,207)MS(C x subj.within groups) + q(4,621)MS(BC x subj. within groups)(r-1)/MS(C x subj. within groups) + MS(BC x subj. within groups) (r-1) = 3.67(0.727) + 3.66(.312)(4-1)/0.727 + (0.312)(4-1) = 3.66

q(0.01) = 4.46(0.727) + 4.44(0.312)(4-1)/0.727 + (0.312)(4-1) = 4.45

Error term for calculating Tukey's HSD for comparisons between quantifiers: = (MS(B x subjects within groups) + MS(BC x subjects within groups)(r-1)/rpn)$^{1/2}$ = (.416/24 x 3)$^{1/2}$ = 0.076

Tukey's HSD = q(0.05) x ERROR = 3.66 X 0.076 = 0.278
q(0.01) x ERROR = 4.45 X 0.076 = 0.338

IF THEN

	B 0.889	E 1.194	T 1.458	N 1.889
B 0.889	-	0.305 *	0.569 **	1.000 **
E 1.194		-	0.264 ns	0.695 **
T 1.458			-	0.431 **

ALL

	B 1.083	T 1.458	E 1.514	N 1.889
B 1.083	-	0.375 **	0.431 **	0.806 **
T 1.458		-	0.056 ns	0.431 **
E 1.514			-	0.375 **

TABLE 4.2f cont.

ANY

	B 1.000	E 1.111	T 1.472	N 1.806
B 1.000	-	0.111 ns	0.472 **	0.806 **
E 1.111		-	0.361 **	0.695 **
T 1.472			-	0.334 *

EVERY

	B 1.167	T 1.389	E 1.583	N 1.889
B 1.167	-	0.222 ns	0.416 **	0.722 **
T 1.389		-	0.194 ns	0.500 **
E 1.583			-	0.306 *

TABLE 4.3

ANALYSIS OF VARIANCE FOR GRADE, QUANTIFER AND MODE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	557.496	287			
BETWEEN	313.246	71			
GRADE	63.465	2	31.730	9.03	0.001
MODE	2.920	1	2.920	0.83	ns
G X M	14.882	2	7.441	2.12	ns
ERROR	231.979	66	3.515		
WITHIN	244.250	216			
QUANTIFIER	25.343	3	8.448	8.05	0.001
G X Q	5.230	6	0.872	0.83	ns
Q X M	2.427	3	0.809	0.77	ns
G X Q X M	3.479	6	0.580	0.55	ns
ERROR	207.771	198	1.049		

TABLE 4.4

ANALYSIS OF VARIANCE FOR GRADE, QUANTIFER AND CONTENT

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	557.496	287			
BETWEEN	313.246	71			
GRADE	63.465	2	31.730	8.97	0.001
CONTENT	0.006	1	0.006	0.00	ns
G X C	16.338	2	8.169	2.31	ns
ERROR	233.437	66	3.573		
WITHIN	244.250	216			
QUANTIFIER	25.343	3	8.448	8.53	0.001
G X Q	5.230	6	0.872	0.88	ns
Q X C	5.007	3	1.669	1.69	ns
G X Q X C	12.690	6	2.115	2.14	0.05
ERROR	195.980	198	0.990		

different content areas with other quantifiers.

CLASSIFICATION OF RESPONSE PATTERNS

As in experiments 2 and 3 data were also classified in terms of response patterns. The response pattern classifications for each quantifier are shown in table 4.5. The predominant classification for subjects at all grades and on all quantifiers was inconsistent. Of the consistent response patterns the majority at P1 were conjunctive; for P3 subjects there were around the same number of conjunctive and inclusion interpretations and by P5 the majority of consistent response patterns were class inclusion/implication interpretations. Comparisons of response pattern distributions between P1 and P3, $\chi^2(3)=14.872$, p 0.01, P3 and P5, $\chi^2(3)= 7.83$, p 0.05, and between P1 and P5, $\chi^2(3)=20.496$, $p<0.001$, were all significant. The distribution of response patterns was similar for all the quantifiers.

DISCUSSION

The results of experiment 4 showed that, as in experiment 2, primary school subjects found it significantly easier to verify universally quantified statements expressed with the quantifier 'all' (74% correct) than with the general conditional 'if then'(68% correct). Significant F ratios on tests of simple main effects of quantifier for B and E type statements only, showed that differences in performance between quantifiers were due to differences in performance on these statement types. Post hoc comparisons between quantifiers for statement types B and E showed that the comparison between 'if then' and 'all' was significant only for E type statements.

TABLE 4.5

CLASSIFICATION OF RESPONSE PATTERNS

EVERY

GRADE	CONJ.	BICOND.	COND.	INCONSISTENT
P1	1	1	0	22
P3	4	0	6	14
P5	2	0	4	18
TOT	7	1	10	54

ALL

GRADE	CONJ.	BICOND.	COND.	INCONSISTENT
P1	4	3	2	15
P3	3	0	4	17
P5	0	0	6	18
TOT	7	3	12	50

IF THEN

GRADE	CONJ.	BICOND.	COND.	INCONSISTENT
P1	5	1	0	18
P3	4	0	3	17
P5	2	1	6	15
TOT	11	2	9	50

ANY

GRADE	CONJ.	BICOND.	COND.	INCONSISTENT
P1	5	1	1	17
P3	7	0	2	15
P5	2	0	3	19
TOT	14	1	6	51

TOTAL

GRADE	CONJ.	BICOND.	COND.	INCONSISTENT
P1	15	6	3	72
P3	18	0	15	63
P5	6	1	19	70
TOT	39	7	37	205

Post hoc comparisons between mean correct response on different quantifiers (Table 4.2b) showed that performance on quantifiers 'all' (74% correct) and 'every' (75% correct) was not significantly different and neither was the comparison between performance on 'if then' (68% correct) and 'any' (67% correct) but all other comparisons were significant. In addition comparisons of mean correct response between 'all' and 'every' and between 'any' and 'if then' for statement types E and B were not significant (Table 4.2d). These results suggest that 'any' is interpreted in a similar way to 'if then' by 6 to 10 year olds while 'every' is interpreted in the same way as 'all', at least in this experiment.

COMPARISON WITH PREVIOUS EXPERIMENTS

For the purposes of comparing the results of experiment 4 with the results of the primary children in experiment 2, correct response on different statement types by primary subjects in experiment 2 for quantifiers 'all' and 'if then' are shown in Table 4.6. Although primary subjects in experiment 2 were from primaries 2, 4 and 6 they were only on average 6 months older than the primary 1, 3 and 5 subjects in experiment 4.

Although in experiment 4 verification of class statements with the quantifier 'all' (74% correct) was still significantly better than that on the general conditional (68% correct) the difference between the levels of correct response on class and general conditional statements in experiment 4 was much smaller than in experiment 2 where 81% of primary subjects' responses on 'all' and 58% of responses on 'if then' were correct. The predicted improvement in performance on conditional statements by presentation of quantifiers in a within subjects design was found - 68% of responses in experiment 4 were correct compared with only 58% in experiment 2. This difference in correct response between experiments 2 and 4 was significant on a test for the difference between two proportions, $Z=2.254$, $p<0.05$. The superior performance of primary children in verifying class statements compared with general conditionals in experiment 2 was entirely due to difficulty with B and E type conditional statements. Performance on both E and B type conditional statements in experiment 4 was significantly better than that in experiment 2 according to tests for the difference between two proportions; for E type statements $Z=3.549$, $p<0.05$, and for B type statements $Z=2.591$, $p<0.05$. The higher level of correct response on both these statement types compared with experiment 2 indicated that presentation of the task in a within subjects design did seem to alert subjects to the interpretation of the general conditional as a universal quantifier. The significantly higher level of correct response to E type conditional statements compared with B type conditional statements in experiment 4 contrasted with the similar levels of correct response on these statement types in experiment 2 and was a further indication that subjects were adopting a more sophisticated interpretation of the general conditional. Despite the predominance of the inconsistent response the response pattern analysis also shows a higher incidence of correct response with 9 inclusion/implication response patterns in experiment 4 compared with only 1 for primary subjects in experiment 2. Although the lower level of correct response on T type statements in experiment 4 compared with experiment 2 was not significant, $Z=1.688$, it was a further indication that subjects attempted to interpret the general conditional as an inclusion/implication relation rather than a conjunction since while conjunctive interpretation produces good

TABLE 4.6

MEAN PERCENT CORRECT RESPONSE AS A FUNCTION OF GRADE, LINGUISTIC FORM AND STATEMENT TYPE FOR PRIMARY SUBJECTS ON EXPERIMENT 2

IF THEN

	P2	P4	P6	TOT
T	88	91	78	86
B	22	22	19	21
E	22	28	31	27
N	100	97	97	98
TOT	58	60	56	58

ALL

	P2	P4	P6	TOT
T	94	78	88	87
B	63	56	72	64
E	72	72	88	77
N	94	100	94	96
TOT	81	77	86	81

performance on T type statements the asymmetry of inclusion leads to errors on T type statements. The relatively poor performance on B type statements was a further indication of the difficulty in dealing with the asymmetry of inclusion. The higher error rate on B type over T type statements indicates that subjects are more likely to evaluate a B type (superset) statement as true than they are to evaluate a T type (subset) statement as false. Although problems in dealing with the asymmetry of the inclusion/implication relation would lead to a similar error rate on B and T type statements there are still some "conjunctive" errors on B type statements. In addition the lower probability of a 'true' statement compared with a 'false' statement possibly produced a bias to evaluating B type statements as true.

While performance on the general conditional in experiment 4 was better than that in experiment 2, performance in verifying statements with the universal quantifier 'all' was worse in experiment 4 (74% correct) compared with experiment 2 (81% correct) although the difference did not quite reach significance, $Z=1.80$, $p<0.1$. The statement type data showed that the trend towards poorer performance on class statements in experiment 4 was caused by difficulty with the asymmetry of the inclusion relation as the poorer performance in verifying only T and B type statements indicated; however the differences in performance on T and B type statements between the two studies was not significant. Smith (1979) reported that children from 4 to 7 years who were required to verify a block of questions involving the existential quantifier prior to verifying a block of universally quantified questions performed substantially worse than a control group who answered the universally quantified statements first. Although no existentially quantified statements were included in experiment 4 as we have seen 'any' and 'if then' are liable to misinterpretation as such and probably the presentation of the quantifiers in the within subjects design of experiment 4 caused the subjects a degree of confusion in dealing with the asymmetry with the universal quantifiers 'all' and 'every'.

Apart from the small decrease in correct response on T and B type statements the results of experiment 4 are similar to those of experiment 2. While there was no difference in correct response on B and T type class statements by 6 and 11 year olds in experiment 1 performance on T type statements expressed with the linguistic quantifier "all" in experiment 4 was significantly better than performance on B type statements. It is interesting to note that these results, like those of experiments 2 and 3, are contrary to Bucci's prediction of an increase in correct response on T type statements with increasing age and no difference in performance on B type or E type statements across grade. A possible reason for the lower error rate on T type over B type statements was the lower incidence of true statements compared with false statements and a consequent trend to evaluating B (superset) relations as true.

The non-significant comparison between E type statements for 'any' and'if then' and the significant differences between these two quantifiers and 'all' and 'every' indicated that 'any' and 'if then' are significantly more likely to be interpreted as conjunction. Although it was thought that 'every' might also be liable to interpretation as conjunction performance on 'every' was more like that on 'all'.

The improvement in performance in verifying the general conditional in experiment 4 was due to the realisation that the general conditional corresponded to class inclusion. Presumably this realisation enabled the subjects to map the general conditional onto the concrete operational classification operations although the significant difference in performance between verification of "all" statements and "if then" statements indicated that subjects still had difficulty with this mapping.

According to Piaget the ability of subjects at the stage of concrete operations to verify class statements like "All A are B" was attributable to their comprehension of the logic of inclusion. If subjects at concrete operations understand the logic of inclusion they should be able to understand the logical consequences of inclusion. The fact that performance on the verification tasks of experiments 1, 2, 3 and 4 was influenced by content (experiment 1) and factors associated with the presentation of the tasks including changes in the stimulus and response probabilities, indicated that logical considerations alone did not account for performance on these tasks. If correct response on the class verification task is mediated by understanding of the logic of inclusion performance by concrete subjects on other tasks requiring an understanding of the logic of inclusion should also be good. In Chapters 5 and 6 two inference tasks - the evaluation and syllogistic reasoning tasks - which require subjects to work out the logical consequences of the inclusion/implication relation are examined.

SUMMARY

Experiment 4 was an extension of the sentence verification task of experiment 2 to include universal quantifiers 'every' and 'all'. The univerally quantified expressions were presented in a within subjects design. Improvement in performance on E and B type general conditional statements indicated that at least part of the difficulty that primary school subjects had in interpreting conditional statements was due to the fact that subjects did not understand that the general conditional corresponds to universal quantification. Presenting the conditional verification task along with other universally quantified expressions in a within subjects design significantly facilitated this understanding although performance on the conditional verification task was still not as good as that on the class verification task. Universal quantifiers 'every' and 'any' were interpreted as similar in meaning to 'all' and 'if then ' respectively.

CHAPTER 5

EVALUATION OF CONDITIONAL AND UNIVERSALLY QUANTIFIED SENTENCES

Another paradigm which purports to test how the conditional is comprehended is the conditional evaluation task. Johnson-Laird and Tagart (1969) developed this task as a means of determining under what circumstances a conditional is judged as true. In the evaluation task subjects were presented with a conditional sentence like:

(5.1) "If there is an A on the left side of the card, then there is a 3 on the right side."

They were then presented with a set of test cards and asked to decide for each card whether it indicated that the conditional was true or false or was irrelevant to the truth of the card. Four different test cards were used like those in Figure 5.1a. The cards were constructed as exemplars of the four different combinations of truth value of the antecedent and consequent of the conditional.

Subjects interpreting sentence 5.1a as a propositional conditional would have classified only P-Q cards as indicating that the sentence was false, with PQ, -PQ and -P-Q classified as indicating the truth of the sentence (see Table 5.1). A biconditional interpretation of the statement would have led subjects to classify PQ and -P-Q as indicating the truth and -PQ and P-Q as indicating the falsity of the statement.

Johnson-Laird and Tagart found that only 4% of their adult subjects interpreted the conditional as a conditional while most subjects adopted what they called a partial or defective truth table, classifying PQ cards as indicating the truth of the statement, P-Q its

Figure 5.1

Cards used in Johnson-Laird and Tagart's conditional evaluation paradigm for sentence:

"If there is an A on the left side of the card there is a 3 on the right side."

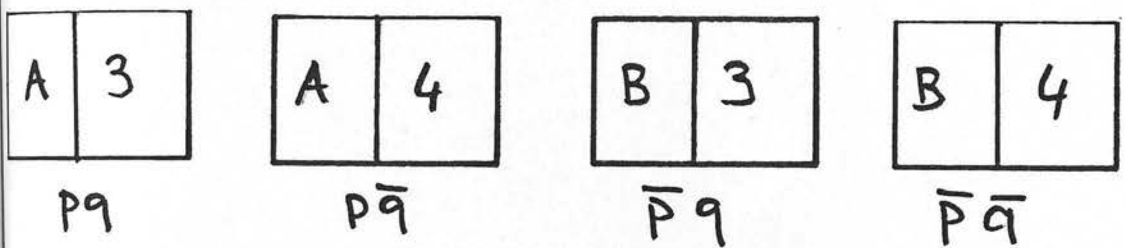


Figure 5.2

Cards used in Ward's propositional reasoning task.

p: Butch robbed the bank.

q: Slim robbed the bank.

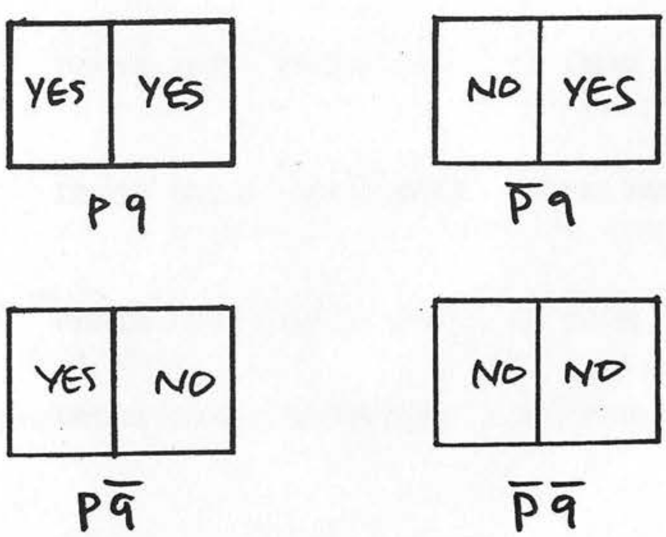


TABLE 5.1a

RESPONSE PATTERNS FOR PROPOSITIONAL INTERPRETATIONS

EXEMPLARS	TRUTH TABLE INTERPRETATIONS			
	CONJ.	BICOND.	COND.	PARTIAL/DEFECTIVE
PQ	T	T	T	T
P-Q	F	F	F	F
-PQ	F	F	T	IRRELEVANT
-P-Q	F	T	T	IRRELEVANT

TABLE 5.1b

RESPONSE PATTERNS FOR GENERAL RULES

EXEMPLARS	PROPOSITIONAL INTERPRETATIONS			
	CONJ.	BICOND.	COND.	PARTIAL/DEFECTIVE
PQ	PROVES TRUE	PROVES TRUE	NO PROOF	PROVES TRUE
P-Q	PROVES FALSE	PROVES FALSE	PROVES FALSE	PROVES FALSE
-PQ	PROVES FALSE	PROVES FALSE	NO PROOF	NO PROOF
-P-Q	PROVES FALSE	PROVES TRUE	NO PROOF	NO PROOF

falsity and classifying -PQ and -P-Q as irrelevant to the truth of the statement (Table 5.1).

Given the poor performance of adults on the conditional evaluation task it would be surprising if children performed well on a task of this kind. Peel(1967), Ward (1972), Paris (1973), Shine and Walsh (1971) and Sternberg (1979) report developmental studies using a task similar to the evaluation task to assess the development of comprehension of various logical connectives including the conditional.

Ward (1972) reports a study of the development of the ability of children ages 7 to 14 years to perform a propositional reasoning task called "Butch and Slim" which was being developed as an item for inclusion in the Operational Thinking Subscale of the British Intelligence Scale (B.I.S). The task consisted of four cards each of which depicted two criminals - "Butch" and "Slim". Under the picture of each criminal on each card was written either "YES, I did rob the bank" or "NO, I did not rob the bank". The four cards comprised the four different combinations of Butch and Slim's affirmations and denials concerning the bank robbery (see Figure 5.2). The subject was then presented with various statements which Butch might have made to the police concerning the roles of Butch and Slim in the bank robbery. There were sixteen different statements based on the sixteen possible combinations of Butch and Slim's assertions e.g. the statement "Slim and I robbed the bank together" was based on propositional conjunction, (p.q). There were four possible implication relations corresponding to $p \supset q$, $-p \supset -q$, $-p \supset q$ and $p \supset -q$. These were expressed using the conditional connective e.g. "If Slim robbed the bank, I robbed the bank." ($p \supset q$). The subject's task was to say whether "Butch and Slim's answers on the cards could be true or not. You are looking for cards which could agree with what Butch said." The test was intended as a test of the development of the understanding of the sixteen combinations of propositional logic which Piaget claimed were acquired at the stage of formal operations.

The test was found to be useful as a developmental test since some statements were easier to verify than others and some items showed clear developmental trends e.g. conjunction was easy for most subjects at all ages tested (8-14 years) but performance on disjunction improved with increasing age. Performance on implication items was very poor with 0% of 8 year olds and still only 17% of 14 year olds correct on these items. These results are in general agreement with the results of several other developmental studies showing that conjunctive statements are understood even by the youngest groups studied (Suppes and Feldman, 1971; Paris, 1973; Neimark and Slotnick, 1970; Sternberg, 1979). These studies also showed that comprehension of disjunction improved with increasing age. Sternberg found an interesting crossover in interpretation of disjunction with younger subjects (7.5-9.5 years) tending to use an inclusive interpretation and older subjects tending to adopt an exclusive interpretation of disjunction. Implication is difficult to understand at all ages but is seldom understood before 12-13 years, Shine and Walsh (1971), Paris (1973), Sternberg (1979).

Although Ward did not discuss the specific nature of errors on the implication items he indicated that subjects seemed to ignore the word "if" and treat the sentence as a conjunction saying that only the card corresponding to the truth of the antecedent and consequent was true. We can assume from this that most subjects classified only PQ as true of the conditional.

Paris (1973) also reports very poor comprehension of the conditional by subjects from 7 years up to 19 years in a kind of evaluation task. Subjects were required to say whether a conditional sentence such as:

(5.2) "If the bird is in the nest then the shoe is on the foot."

was a true or a false description of a picture. The pictures corresponded to the four different exemplars of the combinations of the truth and falsity of the antecedent and consequent of the conditional; for instance the picture corresponding to the truth of the antecedent and the falsity of the consequent would depict "a bird in a nest" and "a shoe which was not on a foot".

Paris reports correct response by subjects on particular exemplars rather than response patterns of individual subjects. He found that subjects make few errors on PQ exemplars (only 6% error rate overall) and P-Q exemplars (only 4.5% error rate overall) but that 93% of responses to -PQ exemplars and 58.5% of responses to -P-Q exemplars were incorrect. The high error rate on -PQ exemplars indicates that virtually no subjects correctly interpreted the conditional as a conditional. The older subjects were, if anything, more consistently incorrect with 100% of 13 and 16 year olds and 95% of 19 year olds responding incorrectly to the -PQ exemplars compared with 75% of 7 year olds. Although the high error rates on -P-Q exemplars (95% errors) and -PQ exemplars (95%errors) indicated that most 7 year olds adopted a conjunctive interpretation of the conditional, there was an increase in correct response on -P-Q with increasing age up to 13 years (57.5% errors) consistent with an increasing tendency to respond biconditionally. There was no further increase in correct response by

16 and 19 year olds.

Because children readily accept arbitrary rules in the context of games, Peel (1967) hit upon the idea of testing comprehension of statements containing logical connectives in this context. The child and the experimenter played a game in which the experimenter had a collection of coloured beads (red, yellow, blue and green) and the child had a collection of counters of the same colours. The child had to pick a counter to go into a box following the experimenter's choice of a bead according to a rule e.g. "If and whenever I draw a red bead, you must also draw a red counter.". Peel looked at childrens' comprehension of three different rules - implication, incompatibility and disjunction.

The game was played both with the experimenter stating the rule and the child making the appropriate response and with the child stating the rule and judging whether the experimenter's choice was appropriate. In the first condition with the implication rule most subjects (ages 5 to 11) drew a red counter (r) when a red bead (R) was drawn (Rr) and drew a non-red counter (-r) when a non-red bead (-R) was drawn (-R-r). Peel understood from this that the children understood the rule as a biconditional.

If this response pattern was adopted in the second condition subjects would accept the experimenter's Rr response and -R-r and reject the R-r response (non-red counter following red bead) and the -Rr response (red counter following non-red bead). Although this was indeed the predominant response pattern for 5 year olds there was a developmental trend observed in acceptance of -Rr responses wih 24% of 5 year olds, 58% of 8 year olds and 82% of 11 year olds correctly accepting a red counter following a non-red bead. This response pattern (the acceptance of Rr, -R-r and -Rr) is the pattern which would be expected from an implication interpretation of the rule. Although children do not initiate an -Rr response, presumably because they do not want to go beyond the rules of the game, over half the 8 year olds and the majority of 11 year olds judged this response as acceptable when produced by the experimenter.

While 13 and 14 year olds in Ward's study, 13, 16 and 19 year olds in Paris's study and adults in Johnson-Laird and Tagart's study apparently fail to understand implication, 58% of Peel's 8 year olds and 82% of 11 year olds accepted responses consistent with an implication interpretation of the conditional and most 5 year olds accepted responses consistent with a biconditional interpretation of the conditional.

The comparison between performance on Peel's task and Paris's, Ward's and Johnson-Laird's tasks however is unrealistic. Although all these tasks purport to study comprehension of implication it is evident that it is not only the subject's understanding or misunderstanding of the implication relation which determines performance on the task: rather factors associated with the structure, presentation and content of the task very largely determine response on the task. These factors can be categorised in terms of those relating to:

(a) Presuppositions concerning the nature of the task which constrain the interpretation of the rule.

(b) The response required and the nature of the conditional.

(c) The content of the task.

(d) Concrete/verbal presentation of the task.

(a) The majority of subjects in Paris's task were inferred to interpret the conditional as a conjunction. A similar interpretation of the conditional rule in Peel's task would mean that subjects would classify only Rr responses as acceptable. However the game was set up in such a way that a conjunctive response pattern was extremely unlikely. The subject was told prior to the game in Peel's study that "we may each pick any colour we like". Given this statement the subject is very unlikely to say that any move by the experimenter in which he chooses a non-red bead (i.e a -R move) is unacceptable. In other words the presuppositions of the task make a conjunctive response unlikely.

(b) From the responses that the majority of Johnson-Laird and Tagart's subjects gave on the conditional evaluation task - the partial truth table responses - it is clear that most subjects did not construe the conditional as true when its antecedent was false. In other words the subjects did not construe the conditional truth-functionally. This however is possibly not as surprising as Johnson-Laird and Tagart seem to think since their "propositional conditional" is really a general conditional (Quine 1944/80) , a general rule concerning a number of cards rather than a propositional conditional which would concern only one card. That the conditional is intended as a general rule is evident from Johnson-Laird and Tagart's assertion that the conditional statement (5.1) is synonomous with (5.2)

(5.2) "There is never a letter 'A' on the lefthand side without there being a number '3' on the right hand side."

The correct response pattern for exemplars in the evaluation task involving a general conditional is different to that for a material conditional. It is never possible to verify a general conditional conclusively but only to falsify it. Only a P-Q exemplar can falsify a general conditional and the other exemplars, PQ, -PQ and -P-Q, offer no proof of the rule (See Table 5.1b). If subjects understand that the rule cannot be verified then the "irrelevant" responses to -PQ and -P-Q exemplars are probably the most appropriate responses they could give, given the choice between "True", "False" and "Irrelevant". "True" responses to PQ exemplars are incorrect if the statement is understood as a general rule. It has been established that subjects do erroneously regard PQ exemplars as verifying a general rule, (Moshman, 1979, O'Brien and Overton, 1982)

In order to judge whether a subject is responding correctly in an evaluation task it is necessary to be clear about whether the statement is a propositional (material) conditional or a general conditional. If the evaluation task concerned a general conditional and subjects adopted the "partial" response pattern of Table 5.1b they would be regarded as responding incorrectly to PQ exemplars but correctly to -PQ and-P-Q exmplars whereas if they adopted the same response pattern with a propositional conditional they would be regarded as responding correctly to PQ exemplars and incorrectly to -PQ and -P-Q exemplars.

Another related point here concerns whether the exemplars are presented for evaluation simultaneously or consecutively. In the original task of Johnson-Laird and Tagart exemplars were presented consecutively. Although the statement used was a general conditional it could be construed as a propositional conditional which was to be evaluated afresh for each card (exemplar). In some studies involving tasks similar to the evaluation task (Ward 1972; Shine and Walsh, 1971; Neimark and Slotnick, 1970) sets of exemplars were presented for evaluation simultaneously. Shine and Walsh for example presented their subjects with a statement for evaluation and a picture of several shapes and instructed the subjects to "cross out figures for which the statement is true". The subject effectively has to consider each shape with respect to the statement just as he has to consider each exemplar with respect to the rule in the evaluation task. However it seems likely that presentation of exemplars simultaneously will increase the liklihood of subjects focusing their attention on the most salient exemplars at the expense of the others and the most salient exemplars will be those mentioned in the rule. Response on the evaluation task is thus likely to be influenced by whether exemplars are presented consecutively or simultaneously.

Looking at the evaluation task in this way brings out the similarity between the evaluation task and the verification task. In the verification task the conditional has to be judged as true or false with respect to a set of exemplars. In the evaluation task the conditional is judged as true or false with respect to one exemplar at a time: the evaluation task is not the same as the verification task even in the case where the exemplars are presented at the same time since even in this case the evaluation of the truth or falsity of the rule is made for each exemplar individually.

In Chapter 1 it was mentioned that according to Ennis's interpretation of Piagetian theory a propositional combination can be verified only if all elements relevant to that combination can be shown to be present and all other elements shown not to be present. Thus an implication can be verified only if PQ, -PQ and -P-Q elements are present and P-Q elements absent. It is possible to verify an implication in terms of Piaget's theory but it would not be possible to do so within the context of the evaluation task since the truth of an implication can only be established by consideration of all exemplars simultaneously. In the evaluation task the subject is required to say whether one element at a time makes the conditional true or false. Even when exemplars are presented simultaneously they are presumed to be judged independently. In order to say whether an implication is true or false in terms of Piaget's theory a judgement would have to be made about the set of exemplars. This wás the judgement required in the conditional verification task.

(c) Another factor which affects response on the evaluation task is the content of the statement to be evaluated. It seems likely that exemplars are more likely to be evaluated correctly when the empirical relationship between the content of the antecedent and consequent is obviously an implication relationship. Marcus and Rips (1979) showed a significantly higher number of biconditional response patterns in both an evaluation task and a syllogistic reasoning task when the content of the problem concerned an empirical situation in which the relationship between the event described in the antecedent and that in the consequent was likely to be understood as a biconditional relationship rather than a conditional relationship. The "biconditional" problem concerned an apparatus in which a ball rolled into a hole and lit a coloured light. Subjects tended to interpret a statement like: "If the ball rolls left the red light flashes" as a biconditional (i.e. as false in the case were the ball does not roll left and the red light flashes as well as in the case where the ball rolls left and the light does not flash) more often than a statement where the relationship between the antecedent and consequent was not likely to be interpreted as a biconditional e.g. "If a fish is striped it is red." or a statement, like that used by Johnson-Laird and Tagart, concerning the relationship between letters and numbers on cards. The empirical relationships between the antecedent and consequent in the latter two problems is what might be termed contingent but plausible. Although many researchers into reasoning with natural language premises regard the effect of the content of the premises on the inferences made as a factor to be examined in a complete account of reasoning with natural language premises, contingent but plausible premises seem to be the ideal to test comprehension of logical relationships since it is as close to "neutral" as can be achieved. In Paris's task the content used was contingent but not very plausible. Most people would find it difficult to imagine how the fact that "a bird is in the tree" is related to the fact that "the shoe is on the foot".

(d) Another factor relating to the presentation of the evaluation task which may influence subjects' responses on the task is whether the data for evaluation are presented linguistically or non-linguistically. In Johnson-Laird and Tagart's original conditional evaluation experiment the data were presented non-linguistically: cards were presented rather than descriptions of cards. Marcus and Rips (1979) on the other hand presented their subjects with verbal descriptions of the relevant attributes of the exemplars to be evaluated although they had also shown their subjects the relevant task materials in the initial description of the task. Marcus and Rips found little difference between the response patterns of their subjects with linguistically presented problems and those of Johnson-Laird and Tagart with non-linguistic presentation: 81% of responses were consistent with a conditional interpretation compared with 83.3% of conditional or partial truth table responses (considered as equivalent to the conditional) in Johnson-Laird and Tagart's task.

It is not clear however that children would find linguistic and non-linguistic information equally easy to deal with. In experiment 5 exemplars were presented verbally since performance on the evaluation task was to be compared with that on the syllogistic reasoning task.

The studies discussed so far have looked at how adults (Johnson-Laird and Tagart, Marcus and Rips) and children (Paris, Peel, Ward, Sternberg) comprehend conditional statements in the evaluation paradigm. There have been few studies concerned with the evaluation of universally quantified statements. Given the logical equivalence of the general conditional and universally quantified statements this omission is surprising.

One of the few studies using the evaluation paradigm to assess how universally quantified statements are understood is a developmental study by Moshman (1979). Moshman presented his subjects with a hypothesis or general rule of the form "If p then q", "If p then not q", "All P are Q" and "No P are Q". Subjects were then presented with eight different data descriptions pertinent to the rule and they were required to say for each data description whether it proved the rule true or false or offered no proof of the rule. The correct implication interpretation of course was that p-q exemplars proved the rule false while pq, -pq and -p-q exemplars offered no proof of the rule.

Moshman looked at the response patterns subjects gave in their evaluations of the data rather than reporting correct responses on specific exemplars. He found an increase in correct implication interpretation with increasing age from 18.75% correct at grade 7 (12 years 10 months) to 42.75% correct at grade 10 (15 years 9 months) to 60.5% correct by college students (19 years 4 months). This increase in correct conditional interpretation was accompanied by a decrease in biconditional and inconsistent response. In contrast to other studies using the evaluation (or hypothesis testing) task with children, Moshman did not identify any conjunctive interpretations of the general rule. However this was probably due to the inclusion in his task of three response categories rather than the two used in other studies. Logically a subject adopting a conjunctive interpretation would evaluate pq instances as proving the rule true and -p-q, -pq qnd p-q as proving the rule false. Inclusion of a third response category probably introduces complexity into the task and makes subjects unsure as to whether "proves false" or "offers no proof" is the correct response on -p-q, -pq and p-q exemplars. Many subjects who respond according to a conjunctive interpretation with two response categories would probably be classified as responding inconsistently with three response categories.

The significant effect of logical form which Moshman found was entirely attributable to the greater number of implication interpretations for negative rules, "No P are Q" and "If p then not-q", than for affirmative rules, "All P are Q" and "If p then q". Subjects apparently found it no easier to evaluate exemplars with respect to the universally quantified statement "All P are Q" than to evaluate exemplars with re:ect to the conditional "If p then q".

Given the large difference in response on class and conditional statements in the verification task of experiment 2 the absence of a significant effect of linguistic form in Moshman's task is surprising since both the verification task and the evaluation task (hypothesis testing task) purport to test comprehension of implication. However the main effect of linguistic form in the verification task was attributable to the difference in performance by subjects up to first year secondary. Since the mean age of Moshman's youngest group was 12 years 10 months it is possible that a younger group of subjects in an evaluation/hypothesis testing task like Moshman's would have shown a significant effect of linguistic form.

An alternative explanation of the absence of an effect of linguistic form is that the evaluation task using both class and conditional statements requires formal operational thinking. Although the good performance of primary school subjects with class statements in the verification task might be taken as an indication that subjects at this age can understand logical class inclusion and should therefore be able to respond correctly on any task involving the comprehension of class inclusion, it seems likely that subjects will be able to solve extensional tasks like the verification task at a younger age than purely linguistic tasks like the evaluation task. In a task like the evaluation task the subject effectively has to use the structure of the sentence to work out the kinds of empirical elements which are compatible with that sentence whereas in the verification task the elements are physically present, or if the sentence is false, any one falsifying instance is sufficient to falsify the sentence.

In experiment 5 subjects from 6 to 17 years were presented with an evaluation task involving the evaluation of pq, p-q, -pq and -p-q exemplars with respect to either a class or conditional general rule. The suitability of the consistency/inconsistency response requirement rather than a proves true/proves false/no proof response requirement has been discussed. However this kind of judgement is obviously not very useful in presenting the task to children since they are unlikely to understand the terms consistent/inconsistent or compatible/incompatible. The task was presented by describing to the child a certain domain e.g. 'insects in a garden' for which it could be said that a certain rule was true e.g. "In my garden all the big insects are striped.". The subject's task was to say whether exemplars with particular attributes, pq, p-q, -pq and -p-q, (insects of a particular size and colour) could belong to this domain. The judgement concerning whether an exemplar is a possible element of a domain is apparently the same type of judgement as the consistency/inconsistency of an exemplar with a general rule but framed in terms that the young child can understand.

EXPERIMENT 5

THE EVALUATION TASK

METHOD

The general procedure and subjects were as described in Chapter 3.

The evaluation task was similar to that developed by Johnson-Laird and Tagart as a means of studying the comprehension of implication. Since the children probably would not understand what words like proof, consistent and compatible mean the task was presented in the manner described below. Two different content vehicles were used in the presentation of the task - insects and shapes. The instructions for the insect problem are presented below (those for the shape problem are presented in Appendix D).

The child was told that the problem was about the insects in Mr Jones' garden and he was read the following story which described the problem and contained the instructions:

"One day when Mr. Jones was in his garden he noticed lots of insects. He watched carefully and saw three different kinds of insect. We don't have a picture of the insects but we do know that Mr. Jones said this about the insects in his garden. He said: "In my garden all the big insects are striped.". We know too that this sentence is true. Remember there are three different kinds of insect in Mr. Jones' garden; Mr. Jones said that in his garden all the big insects are striped and this is true. What I'd like you to do now is to say whether you think Mr. Jones could have insects like this in his garden: "This insect is big and it is striped." So we have here an insect which is big and striped and you have to say whether you think

Mr. Jones could have an insect like this in his garden, YES or NO. Now we have here another insect...etc."

The question was repeated eight times for the eight different kinds of exemplar formed by combining positive and negative values of the attribute mentioned in the rule. These are shown in Table 5.2. There are of course only four different kinds of exemplar (insect) but each was presented twice, the order of presentation of the attributes being different in the two cases.

The story was read aloud to the child and the rule and exemplars were also presented on separate written cards, with the exemplars presented in random order. For half the subjects the rule was a universally quantified statement and for the other half it was a general conditional.

It was emphasised to the subject that there were three different types of insect in Mr. Jones' garden to try and minimise the liklihood that the subject would consider that only exemplars with the attributes mentioned in the rule were relevant. It was also emphasised that he should say whether an insect might be in Mr. Jones garden not whether it definitely was in the garden.

RESULTS

Mean percent correct response to the four different exemplars, pq, p-q, -pq and -p-q, for the two different linguistic forms for subjects at different grades are shown in Table 5.3 and represented graphically in Figures 5.3a and 5.3b. A 6 (grade) X 2 (linguistic form) X 4 (type of exemplar) analysis of variance with grade and linguistic form as between subjects factors and instance type as a within subjects factor

Table 5.2

Rule: In my garden all the big insects are striped

Examples:

(1) This insect is big
and it is striped.

(2) This insect is big
and it is not striped.

(3) This insect is not big
and it is striped.

(4) This insect is not big
and it is not striped.

(5) This insect is striped
and it is big.

(6) This insect is striped
and it is not big.

(7) This insect is not striped
and it is big.

(8) This insect is not striped
and it is not big.

TABLE 5.3

MEAN PERCENT CORRECT RESPONSE AS A FUNCTION OF LINGUISTIC FORM, GRADE AND TYPE OF EXEMPLAR

	PQ			-P-Q			-PQ			P-Q			TOT		
	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT	CL	CO	TOT
P2	81	78	80	28	19	23	34	31	33	75	69	72	55	49	52
P4	84	100	92	16	31	23	34	38	36	72	75	73	52	61	56
P6	100	94	97	16	41	28	19	28	23	78	69	73	53	58	56
S1	100	100	100	75	81	78	38	25	31	100	94	97	78	75	77
S3	100	94	97	100	81	91	78	41	59	88	84	86	91	75	83
S5	100	94	97	88	91	89	81	50	66	94	100	97	91	84	88
TOT	94	93	94	54	57	55	47	36	41	85	82	83	70	67	68

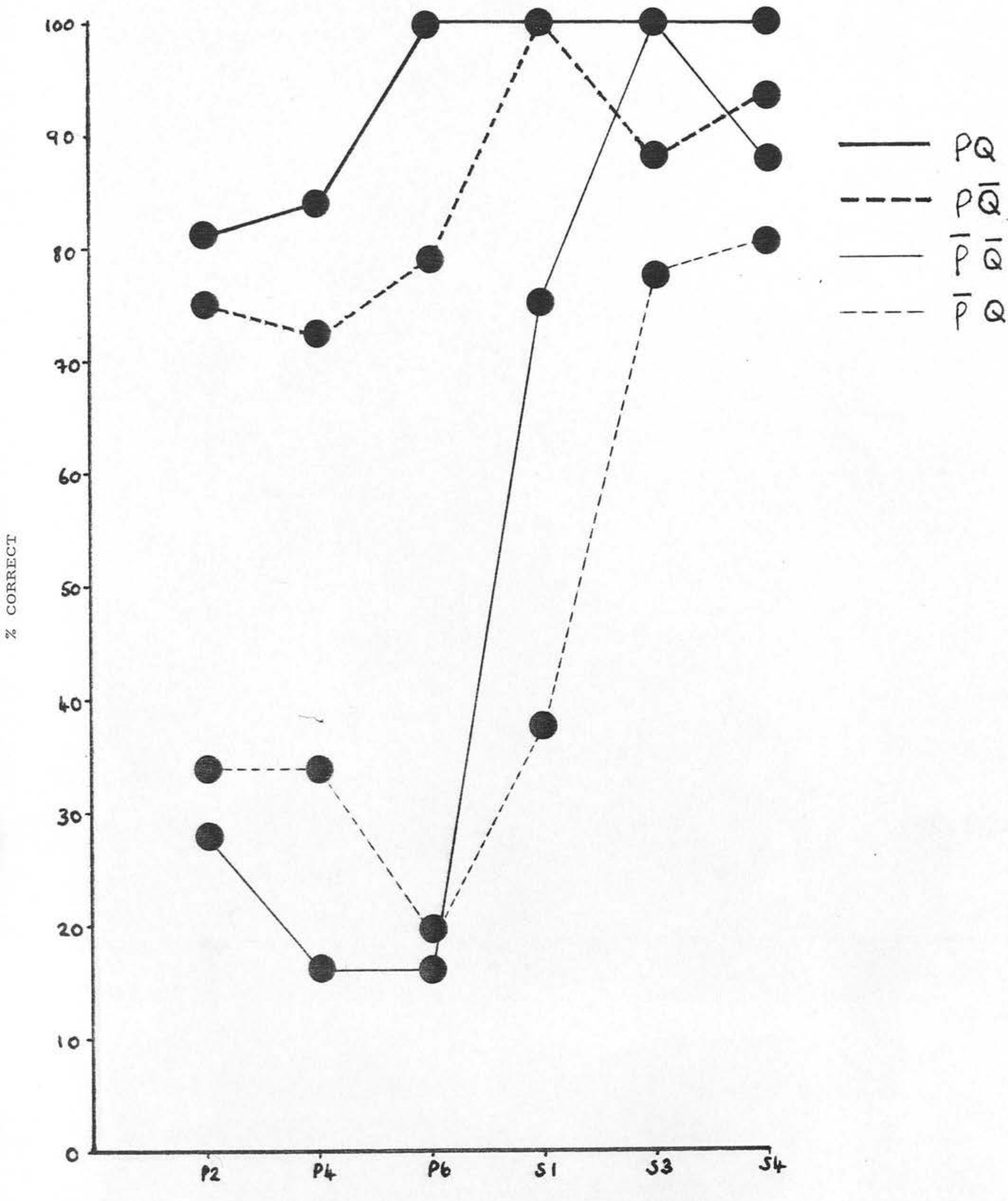


FIGURE 5:3a

PERCENT CORRECT RESPONSES TO DIFFERENT INSTANCES (PQ,P-Q,-P-Q, & -PQ) FOR CLASS RULE PLOTTED AGAINST GRADE

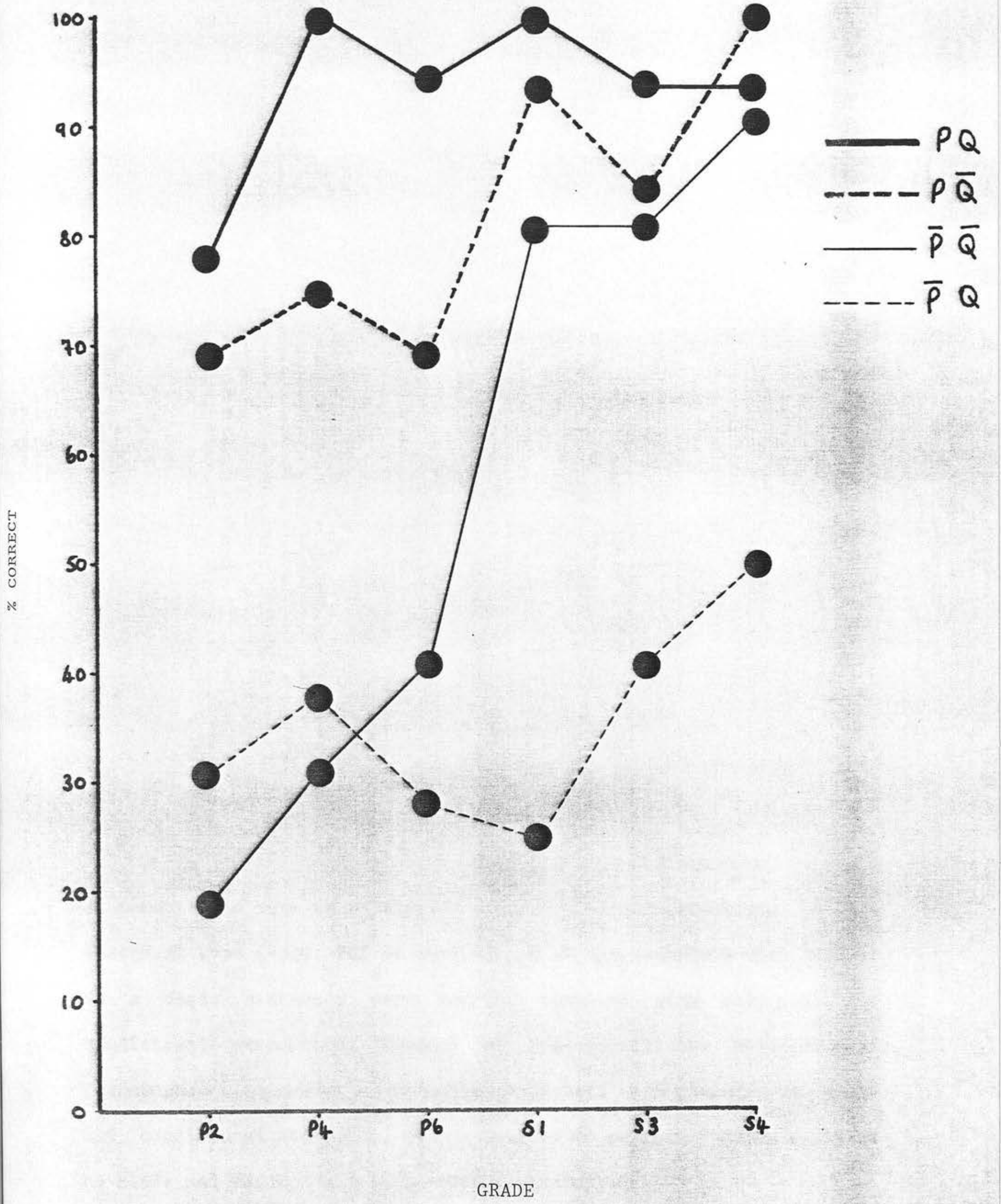


FIGURE 5:3b

PERCENT CORRECT RESPONSES TO DIFFERENT INSTANCES (PQ,P-Q,-P-Q, & -PQ)

FOR CONDITIONAL RULE PLOTTED AGAINST GRADE was carried out on correct response. The results of this analysis are shown in Table 5.4.

There was a highly significant main effect of grade, $F(5,180)=28.49$, $p<0.001$, reflecting mainly the superior performance of the secondary subjects. The main effect of linguistic form was not significant but the grade by linguistic form interaction was significant, $F(5,180)=2.41$, $p<0.05$. The highly significant main effect of exemplar, $F(3,540)=106.40$, $p<0.001$, confirmed that some instances are easier to evaluate than others: specifically 93% of pq, 83% of $p\bar{q}$, 55% of $\bar{p}\bar{q}$ and 41% of $\bar{p}q$ exemplars were correctly evaluated. The highly significant grade by exemplar interaction, $F(15, 540)=6.47$, $p<0.001$, reflected the similar levels of correct response on pq and $p\bar{q}$ exemplars for subjects at all ages in contrast with the increase in correct response on $\bar{p}q$ and $\bar{p}\bar{q}$ exemplars with increasing age.

An analysis of the simple main effects of linguistic form at different grades showed that linguistic form was only significant for subjects at third year secondary, $F(1,180)=7.87$, $p<0.01$ (see Table 5.4a). Although the Grade X Linguistic form X Exemplar type interaction was not significant the superior performance by S3 subjects on class statements did seem to be largely attributable to performance on one statement type ($\bar{p}q$): 78% of evaluations of $\bar{p}q$ exemplars with respect to a class statement were correct compared with only 41% for conditional statements. Because of the significant difference in performance with grade in evaluating exemplars with respect to class and conditional statements it was decided to perform further analyses on class and conditional statements separately.

TABLE 5.4

ANALYSIS OF VARIANCE FOR GRADE, LINGUISTIC FORM AND EXEMPLAR

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	559.72	767			
BETWEEN	147.47	191			
GRADE	62.53	5	12.51	28.49	0.001
LING. FORM	0.69	1	0.69	1.57	ns
GL	5.30	5	1.06	2.41	0.05
ERROR	78.95	180	0.44		
WITHIN	412.25	576			
EXEMPLAR TYPE	134.69	3	44.90	106.40	0.001
EG	40.97	15	2.73	6.47	0.001
EL	2.48	3	0.83	1.95	ns
EGL	6.25	15	0.416	0.99	ns
ERROR	227.87	540	0.422		

TABLE 5.4a

SUMMARY TABLE OF SIMPLE EFFECTS OF LINGUISTIC FORM AT GRADE(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	0.383	1	0.383	0.87	ns
P4	1.125	1	1.125	2.56	ns
P6	0.281	1	0.281	0.64	ns
S1	0.125	1	0.125	0.28	ns
S3	3.445	1	3.445	7.87	0.01
S5	0.633	1	0.633	1.44	ns
ERROR	78.95	180	0.439		

EVLALUATION OF EXEMPLARS WITH RESPECT TO A CLASS RULE

A 6 (grade) X 4 (type of exemplar) analysis of variance with grade as a between-subjects factor and exemplar type as a within-subjects factor was carried out on correct response to the class statement exemplars as a preliminary to calculating simple effects (see Table 5.5). Both factors, grade, $F(5,90)=28.90$, $p<0.001$, and exemplar, $F(3,270)=50.00$, $p<0.001$, and the grade by exemplar interaction, $F(15,270)=4.63$, $p<0.001$, were significant.

Tests of simple interactions were carried out in the first place on adjacent grades and significant differences were found between P6 and S1, $F(3,270)=4.91$, $p<0.005$, and between S1 and S3, $F(3,270)=4.57$, $p<0.005$. No other tests between adjacent grades were significant but all tests of simple interactions between primary and secondary grades were significant (see Table 5.5a).

Tests of simple main effects of type of exemplar revealed significant F ratios for all primary grades and for S1 but not for S3 and S5 (see Table 5.5b) suggesting that performance on this task was mature by third year secondary.

Tests of simple main effects of grade for different exemplars showed that the F ratios for pq and p-q exemplars were not significant while F ratios for -p-q, $F(3,357)=24.37$, $p<0.001$, and -pq exemplars, $F(3,357)=11.19$, $p<0.001$, were significant (see Table 5.5c). This indicates that improvement in performance with increasing age was entirely due to improvement in performance on -pq and -p-q exemplars.

Newman Keuls comparisons of exemplar means revealed significant differences between all exemplars except -p-q and -pq (see Table 5.4d).

TABLE 5.5

ANALYSIS OF VARIANCE FOR CLASS STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
TOTAL	273.039	383			
BETWEEN	75.789	95			
GRADE	46.680	5	9.336	28.90	0.001
ERROR	29.109	90	0.323		
WITHIN	197.250	288			
EXEMPLAR	60.445	3	20.148	50.00	0.001
GE	27.977	15	1.865	4.63	0.001
ERROR	108.828	270	0.403		

TABLE 5.5a

TESTS OF SIMPLE INTERACTIONS BETWEEN GRADES FOR CLASS STATEMENTS: F RATIOS AND SIGNIFICANCE LEVELS

	P2	P4	P6	S1	S3	S5
P2	-	0.362 ns	1.990 ns	2.610 $p<0.05$	5.810 $p<0.001$	3.335 $p<0.05$
P4		-	0.955 ns	4.625 $p<0.005$	8.395 $p<0.001$	5.246 $p<0.005$
P6			-	4.911 $p<0.005$	12.892 $p<0.001$	9.770 $p<0.001$
S1				-	4.569 $p<0.005$	3.928 $p<0.025$
S3					-	0.536 ns

TABLE 5.5b

SUMMARY TABLE OF SIMPLE EFFECTS OF EXEMPLAR FOR CLASS STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
P2	14.313	3	4.771	11.84	0.001
P4	19.688	3	6.563	16.28	0.001
P6	34.625	3	11.542	28.64	0.001
S1	16.750	3	5.583	13.85	0.001
S3	2.172	3	0.724	1.80	ns
S5	1.250	3	0.417	1.03	ns
ERROR	108.828	270	0.403		

TABLE 5.5c

SUMMARY TABLE SIMPLE MAIN EFFECTS OF GRADE FOR CLASS STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
PQ	2.552	5	0.510	1.33	ns
P-Q	4.000	5	0.800	2.09	0.1
-PQ	21.427	5	4.285	11.19	0.001
-P-Q	46.677	5	9.335	24.37	0.001
ERROR	137.937	360	0.383		

TABLE 5.5d

NEWMAN KEULS COMPARISONS BETWEEN EXEMPLARS FOR CLASS STATEMENTS

EXEMPLAR	-PQ 0.948	-P-Q 1.073	P-Q 1.688	PQ 1.885	r	s q(.95)(r,270) s q(.99)(r,270)
-PQ 0.948	-	0.125 ns	0.740 **	0.937 **	4	0.240 0.293
-P-Q 1.073		-	0.615 **	0.812 **	3	0.218 0.273
P-Q 1.688			-	0.197 *	2	0.182 0.241

EVALUATION OF EXEMPLARS WITH RESPECT TO A GENERAL CONDITIONAL

The results of a 6 (grade) X 4 (exemplar type) analysis of variance on correct response on the conditional evaluation task are shown in Table 5.6. Significant F ratios were found for grade, $F(5,90)=7.83$, $p<0.001$, exemplar, $F(3,270)=98.81$, $p<0.001$, and the grade X exemplar interaction, $F(15,270)=2.06$, $p<0.05$.

Tests of simple interactions on adjacent grades revealed a significant F ratio only for the P6/S1 comparison, $F(3,270)=3.38$, $p<0.05$ (see Table 5.6a). Tests on simple interactions for all primary/secondary comparisons were significant but no primary/primary or secondary/secondary comparisons were significant.

Tests of simple main effect of exemplar revealed significant F ratios at all grades (see Table 5.6b). Tests of simple main effects of grade showed that only the F ratios associated with p-φ exemplars, $F(5,360)=2.73$, $p<0.025$, and $-p$-φ exemplars, $F(5,360)=14.78$, $p<0.001$, were significant (see Table 5.6c).

Newman Keuls comparisons between exemplars revealed significant differences between all exemplars (see Table 5.6d).

To establish whether there was an effect of the order of presentation of tasks on response on the evaluation task an analysis of order effects was carried out. Half the subjects responded to the verification tasks first while for the other half the evaluation and syllogism tasks preceded the verification tasks. It was thought possible that the concrete content of the verification tasks would facilitate response on the evaluation and syllogism tasks for those

TABLE 5.6

ANALYSIS OF VARIANCE FOR CONDITIONAL STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
TOTAL	285.990	383			
BETWEEN SUBJ.	70.990	95			
GRADE	21.521	5	4.304	7.83	0.001
ERROR	49.479	90	0.550		
WITHIN SUBJS.	215.000	288			
EXEMPLARS	106.719	3	35.573	98.81	0.001
GE	11.124	15	0.742	2.06	0.05
ERROR	97.157	270	0.360		

TABLE 5.6a

TESTS OF SIMPLE INTERACTIONS BETWEEN GRADES FOR CONDITIONAL STATEMENTS
F RATIOS AND SIGNIFICANCE LEVELS

	P2	P4	P6	S1	S3	S5
P2	-	0.486 ns	1.295 ns	7.483 $p<0.001$	5.420 $p<0.005$	5.960 $p<0.001$
P4		-	0.636 ns	6.568 $p<0.001$	5.468 $p<0.005$	6.794 $p<0.001$
P6			-	3.383 $p<0.05$	2.574 ns	3.842 $p<0.025$
S1				-	1.187 ns	1.468 ns
S3					-	1.244 ns

TABLE 5.6b

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF EXEMPLAR FOR CONDITIONAL STATEMENTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	15.797	3	5.266	14.63	0.001
P4	20.188	3	6.729	18.69	0.001
P6	16.563	3	5.521	15.34	0.001
S1	22.500	3	7.500	20.83	0.001
S3	10.625	3	3.542	9.84	0.001
S5	9.922	3	3.307	9.19	0.001
ERROR	97.157	270	0.360		

TABLE 5. c

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF GRADE FOR CONDITIONAL STATEMENTSTS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
PQ	2.052	5	0.410	1.01	ns
P-Q	5.552	5	1.110	2.73	0.025
-PQ	2.708	5	0.542	1.33	ns
-P-Q	30.083	5	6.017	14.78	0.001
ERROR	146.636	270	0.407		

TABLE 5.6d

NEWMAN KEULS COMPARISONS BETWEEN EXEMPLARS

EXEMPLAR	-PQ 0.708	-P-Q 1.146	P-Q 1.635	PQ 1.865	r	s q(.95)(r,270) s q(.99)(r,270)
-PQ 0.708	-	0.438 **	0.927 **	1.157 **	4	0.225 0.275
-P-Q 1.146		-	0.489 **	0.719 **	3	0.205 0.256
P-Q 1.635			-	0.230 **	2	0.171 0.226

subjects who received the verification tasks first since the verification tasks involved a pictorial representation of elements compatible with the class and conditional rules. Those subjects who responded to the verification tasks with class statements received conditional statements on the inference tasks. Since primary subjects performed much better on class verification than on conditional verification it was possible that any benefits of prior presentation of the verification tasks would be greater when the verification tasks concerned class statements, i.e. when the evaluation task concerned conditional statements. To establish whether there were any effects of presentation order a 6 (grade) X 2 (linguistic form) X order (2) analysis of variance was carried out on correct response. The results are shown in Table 5.7. Only the main effect of grade, $F(5,168)=30.82$, $p<0.001$, and the grade X linguistic form interaction, $F(5,168)=2.87$, $p<0.025$, were significant. The absence of a main effect of presentation order or any interaction of presentation order with other factors indicated that response on the evaluation task was not affected by prior presentation of the verification tasks.

CLASSIFICATION OF RESPONSE PATTERNS

Table 5.8 shows the number of subjects at each grade who responded according to one of the consistent response patterns - conjunctive, biconditional or conditional. There was no difference in the overall distribution of response patterns between class and conditional statements, $\chi^2(3)=6.09$, p 0.2 and consequently analysis on response patterns was performed on the data collapsed across statement form. The number of response patterns classified as consistent increased with increasing age from 19% at P2 to 72% at S5. This increase in consistency with increasing age was significant, $\chi^2(5)=23.63$, $p<0.001$.

TABLE 5.7

ANALYSIS OF VARIANCE FOR GRADE, LINGUISTIC FORM AND ORDER

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	608.828	191			
GRADE	272.297	5	54.459	30.82	0.001
LINGUISTIC FORM	2.755	1	2.755	1.56	ns
ORDER	0.130	1	0.130	0.07	ns
G X LF	25.339	5	5.068	2.87	0.025
G X O	4.714	5	0.943	0.53	ns
LF X O	0.006	1	0.006	0.00	ns
G X LF X O	6.712	5	1.342	0.76	ns
ERROR	296.875	168	1.767		

TABLE 5.8

CLASSIFICATION OF RESPONSE PATTERNS

CLASS RULE

	CONJ.	BICOND.	COND.	INCONSIST.
P2	2	1	0	13
P4	5	0	0	11
P6	8	0	0	8
S1	3	4	3	6
S3	0	2	8	6
S5	1	0	9	6
TOTAL	19	7	20	50

CONDITIONAL RULE

	CONJ.	BICOND.	COND.	INCONSIST.
P2	3	0	0	13
P4	4	2	1	9
P6	5	1	0	10
S1	2	7	2	5
S3	3	3	4	6
S5	1	5	7	3
TOTAL	18	18	14	46

TOTAL

	CONJ.	BICOND.	COND.	INCONSIST.
P2	5	1	0	26
P4	9	2	1	20
P6	13	1	0	18
S1	5	11	5	11
S3	3	5	12	12
S5	2	5	16	9
TOTAL	37	25	34	96

The predominant consistent response pattern for primary school children was a conjunctive/structure neutral interpretation although the majority of primary school response patterns were inconsistent. This changed to a biconditional/equivalence interpretation at S1 and to a conditional interpretation at S3. The only significant adjacent grade comparison was between P6 and S1, $\chi^2(3)=18.58$, $p<0.001$. All non-adjacent comparisons were significant except the P2/P6 comparison, a further indication that improvement in performance only occurred for secondary subjects.

DISCUSSION

The evaluation task is like a hypothesis testing task although as presented here the "hypothesis" is a domain specific rule which is known to be true and the data are presented as exemplars (elements whose attributes have specific relationships with respect to the rule) to be evaluated as possible or not possible elements of the domain. The judgement is thus similar to a compatibility or consistency judgement. Although subjects at all ages seemed to understand what was required of them in the task the younger children did not perform very well. The rule has the effect of restricting the kinds of element which are possible elements of the domain by excluding p-q exemplars. Primary school children, however seemed to think that only those instances with the attributes mentioned in the rule were possible elements of the domain. This interpretation of a class rule leads to what has been called a structure neutral type of response pattern while for the conditional rule a conjunctive response pattern results. The significance of the comparison between P6 and S1 on the tests of simple interactions on the exemplar analysis for both class and conditional rules reflects the improvement in performance between the ages of 10.5 to 12.5 years in accepting as possible elements of the domain exemplars other than those possessing the attributes mentioned in the rule. Overall 78% of -p-q exemplars were accepted by first year secondary subjects compared with only 29% by primary six subjects. There was a corresponding increase in correct rejection of p-q exemplars from 73% at P6 to 97% at S1 (collapsed across class and conditional statements) but subjects at S1 still had difficulty with -pq exemplars with only 31% of the responses by S1 subjects on this exemplar correct compared with 23% of P6 responses. The significant comparison between S1 and S3 on the tests of simple interactions for class statements reflects the increase in acceptance of -pq exemplars between S1 and S3 as well as a further small increase in correct acceptance of -p-q exemplars in contrast with the relatively stable performance on pq and p-q exemplars. By third year secondary there was ~~no~~ significant difference in performance on the four different exemplars for class statements as shown by the non-significance of the F ratio for tests on simple main effects of statement type at S3. Although subjects at S1 realized that exemplars other than pq exemplars were possible elements of the domain this realization appeared before the ability to work out exactly which exemplars were possible. The difficulty S1 subjects still had with the -pq exemplar compared with their good performance in accepting -p-q exemplars indicated that the main difficulty was with the asymmetry of the class statements in allowing -pq but not p-q exemplars. This was confirmed by the classification of response patterns which showed several biconditional interpretations at S1 and fewer conjunctive interpretations, in contrast with the prevalent conjunctive interpretation for primary subjects. The change in the response pattern classification from P6 to S1 was significant. Only the modified classification for the S1/S3 comparison was significant but the S1/S5 comparison for the standard response pattern classification was significant, reflecting a gradual change to the correct response.

Only the P6/S1 comparison for tests of simple interactions was significant for the conditional statement analysis. In addition the F ratios for the tests of simple main effects of statement type for the conditional statement were significant at all grades: the significance of the tests of simple main effects of statement type for S3 and S5 subjects on the conditional statement contrasted with their non-significance on class statements. The difference was mainly due to the superior performance by S3 and S5 subjects in evaluating -pq instances with respect to a class rule as opposed to a conditional rule. The significance of the comparison between -pq and -p-q exemplars on the Newman Keuls comparisons for conditional statements but not for class statements (see Tables 5.5d and 5.6d) was further evidence of the difficulty that subjects had in evaluating -pq with respect to a conditional rule. Although S1 subjects also had problems in evaluating -pq exemplars with the class rule, by S3 subjects could cope with this exemplar. Analysis of simple main effects of grade for conditional statements showed that significant F ratios were found for -p-q and p-q exemplars but not for pq and -pq exemplars. This also contrasted with the class statement analysis which showed significant F ratios for-pq and -p-q exemplars. There was a significant improvement with age in evaluating -p-q exemplars for both class and conditional rules but only for the class rule was the improvement in evaluating -pq significant and only for the conditional rule was the increase in correct rejection of p-q exemplars significant although it was a trend for the class statements.

The results of experiment 5 are compatible with the view that correct evaluation of exemplars with respect to a universally quantified or general conditional statement requires formal operational thinking. The insight that exemplars other than those mentioned in the rule are possible elements of the domain seems to occur around the age of 12 years, the age regarded by Piaget as the advent of formal operational thinking. This insight is very like that described by Piaget as an identifying feature of formal thought:

"There is no doubt that the most distinctive feature of formal thought stems from the role played by statements about possibility relative to statements about empirical reality." (Inhelder and Piaget, 1958, p. 245.)

"The most distinctive feature of formal operational thought is this reversal in direction between reality and possibility." (Inhelder and Piaget, 1958, p. 251)

According to Piaget the concrete operational child is limited to structuring reality and his understanding of language reflects this fundamental limitation. He regards statements such as the class and conditional general rules as "direct mappings" of reality and cannot use the structure of the sentence to work out the different possible elements which are compatible with the rule. The formal operational subject on the other hand can understand the hypothetical or higher order nature of the general rule and can understand that the general rule is compatible with different possible empirical situations.

The concrete operational subject's grasp of additive classification should enable him to understand that "All A are B" entails the possibility of b which are not-a since $A + A' = B$. It might be predicted that this ability should enable the concrete operational subject to understand from the statement "All A are B" that ab and -ab elements are possible elements of a domain for which "All A are B" is true. However as the evaluation task demonstrated subjects did not accept -ab elements as possible domain elements until adolescence. It is not clear whether the concrete operational subject would also be expected to be able to accept -a-b elements as compatible with the inclusion relation, $A < B$, although it seems likely that they would not since the structures of the concrete operational subject are not integrated into a complete system. The evaluation task showed that subjects cannot accept -a-b elements until around adolescence when he understands that elements other than those mentioned in the rule are compatible with it.

Although Piaget talks of the groupings of concrete operations as logical groupings they are not logical in the standard sense of being independent of content and constituting complete systems. The concrete operational subject's comprehension of the inclusion of one class in another is initially based upon an understanding of concrete classes which are neighours in a classification scheme: e.g. dogs + dogs which are not mammals = mammals. Given a statement such as "All dogs are mammals" the concrete subject knows that there may be mammals which are not-dogs. An evaluation task based on comprehension of class inclusion between neighbours in a hierarchical classification scheme would be like this:

"All dogs are mammals. Is it possible that there are things which are:

a) both dogs and mammals (ab)

b) things which are not-dogs but are mammals (-ab)

c) things which are dogs but are not-mammals (a-b)

d) things which are not-dogs and not-mammals (-a-b)?"

Response in a task such as this could be generated independently of the class inclusion statement and would constitute a test of the subject's semantic memory representation, which is an interesting study in its own right as Smith (1979) has shown. Indeed, as mentioned in Chapter 3, Smith showed that even 4 year old children responded well on a task similar to this. However the aim of the evaluation task was to establish to what extent the subject can use the linguistic structure of the class inclusion and implication expressions to work out the inferential properties of these expressions in the absence of semantic (as in the task above) or empirical (as in the verification task) support.

Although comprehension of "All A are B" is said to depend upon an understanding of additive classification, when the inclusion relationship concerns an inclusion relationship which is contingent rather than logically necessary (i.e. an inclusion relationship which is empirically true rather logically necessary) theoretically comprehension of the logical consequences of the linguistic expression of class inclusion seems to depend in a more obvious way on the logical multiplication of classes rather than on their logical addition since it depends upon recognition that class inclusion corresponds with logical implication and logical implication is a propositional operation which can only be generated from the products of the logical multiplication of two classes. Logical multiplication seems to be related to understanding of "All A are B" in the following way: the concrete operational subject can logically multiply (A + A') and (B + B') to generate the products of this logical multiplication, AB + A-B + -AB + -A-B. According to Piaget the concrete subject does not understand the specific combination of elements from classes AB, -AB and -A-B and the absence of A-B elements as corresponding to class inclusion since this combinatorial ability is one of the abilities defined by Piaget as specifically formal operational. Multiplicative classification does not per se lead to understanding of "All A are B" but it is a necessary precursor for understanding "All A are B".

The difficulties in trying to understand exactly what the distinction is between the abilities of the concrete and formal operational subject in understanding class inclusion and implication were mentioned in Chapter 1. Piaget's claim in "Traité de Logique" (1949) that implication corresponds with inclusion $P \subset Q$ and his illustration of implication in terms of class inclusion made the distinction particularly unclear. It seems however that Piaget's contention is that, although class inclusion and implication are logically equivalent, the formal operational subject has a more complex representation of the logical properties of inclusion and implication. Class inclusion can be understood at the concrete level provided the classes are obviously hierachically related or, as in the verification task, the class inclusion statement refers to an empirical inclusion relationship.

One of the most interesting results of the evaluation task was the poor performance by primary school subjects with class statements. The non-significance of the simple effects of linguistic form for the younger children for the grade by linguistic form interaction contrasts with the sizeable difference in correct response on class and conditional statements by primary subjects in the verification task - 53% of class evaluations and 56% of conditional evaluations by primary children were correct compared with 79% of class verifications and 57% of conditional verifications. The significant grade by linguistic form interaction reflected the superior performance of S3 subjects only for the class rule compared with the conditional rule. Comprehension of a universally quantified statement: "All A are B" in an extensional task like the verification task does not guarantee that the statement is understood in the absence of appropriate concrete support. The primary school children apparently cannot use the linguistic structure of the universally quantified statement to work out what the statement entails at least in the operationalization of the evaluation task used here. Although some (e.g. Ennis) would understand the poor performance of the primary school subjects in the evaluation task as evidence against Piaget's claim that concrete operational children can reason in terms of classes it would seem that, on the contrary, the results of the verification and evaluation tasks with class statements are entirely congruent with Piaget's theory in showing that subjects at the stage of concrete operations can deal with the quantification of classification when the problem concerns an extensional situation as in the verification task but when the problem requires the subject to work out logical consequences of the inclusion relationship in the absence of empirical or semantic support the concrete subject has difficulties. The relationship between the terms in the class rule in the evaluation task was contingent but plausible i.e. there was no necessary relationship between the size and markings of insect or between the shape and colour of shapes but it was not implausible that within a specific domain a certain relationship should hold. In order to solve the problem the subject had to rely solely on the logical structure of the class inclusion relationship.

MENTAL MODELS

Johnson-Laird's theory of mental models has been mainly used to explain errors in syllogistic reasoning tasks and spatial inference tasks. If it is intended as a more general theory of inferential ability it should also be able to explain errors on a wide variety of inference tasks including the evaluation task. It is interesting to consider whether the mental model theory can account for the errors in the evaluation task.

Johnson-Laird argues that in making propositional inferences subjects do not need to have knowledge of inference rules because once they have learned the truth conditions for the propositional connectives they can make inferences by considering these truth conditions in conjunction with the conditions under which the other premises are true or false and working out from there what must be true by excluding contingencies which cannot be true. Johnson-Laird underlines the difference between this method of making inferences and making inferences with inference rules. The former method is semantically based relying on knowledge of the meaning of the connective: knowledge of the logical properties of a connective emerge from the subject's comprehension of the meaning of that connective. The latter method relies on knowledge of formal rules of inference. However although it is possible that subjects make propositional inferences by considering truth functional properties of connectives Johnson-Laird argues that this method of inference would apply to a very limited number of inferences. In order to account for the wider variety of natural language inferences which people find acceptable Johnson-Laird proposed the mental model theory as a more general theory of inference. According to Johnson-Laird, in understanding the meaning of a sentence a mental model of the sentence is constructed which corresponds to the states of affairs which the premises describe and this model depends on general knowledge and knowledge of context. The model can be used as a basis for making inferences.

In comprehending a quantified sentence such as "All A are B" the reasoner constructs a mental model of the sentence by representing prototypical exemplars of the terms A and B and the relationship between the terms. The model which the subject constructs consists of an arbitrary number of 'as' and 'bs'; each 'a' is identified with a 'b' but there are possibly some 'bs' which are not 'as' (see Fig. 5.4a). Johnson-Laird claims that mental models can also cope with the minor differences in meaning between natural language quantifiers like 'every' and 'all' which are treated in standard logic as meaning the same. It is not clear how the general conditional would be represented in terms of mental models but if the general conditional is understood as a universal quantifier it would presumably be represented in a similar way to the quantifier "all".

From the model of "All A are B" in Figure 5.4a the subject would readily be able to verify that ab elements were possible elements of the domain and also that -ab elements were possible domain elements. However consideration of the model would apparently lead subjects to deny that -a-b and a-b elements were possible since no such elements are represented in the model. The problem with this is that although a-b elements are not compatible with the statement -a-b elements are

Mental model representations of "All A are B."

Figure 5.4a

a = b

a = b

(b)

Figure 5.4b

a = b

a = b

Euler circle representations of "All A are B."

Figure 5.5a

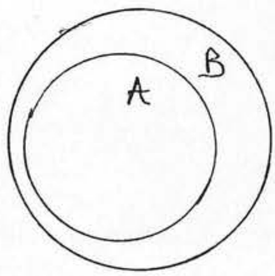
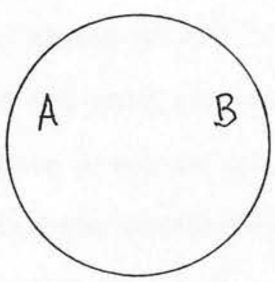


Figure 5.5b



compatible with the statement. Since the mental model theory of inference is primarily intended as a psychological account of inferential ability rather than a theory of how valid logical inferences can be derived this problem might not be important if the model made predictions which were psychologically correct i.e. if subjects did accept only ab and -ab elements as compatible with the model. However this response pattern was not one which was observed in the evaluation task in experiment 5 and it is not one which is typically found in evaluation tasks. The results of experiment 5 showed that the response patterns which any theory would have to account for are the conjunctive/structure neutral, equivalence/biconditional and conditional/class inclusion response patterns.

The mental model theory can readily account for the conjunctive/structure neutral response pattern on the evaluation task if it is assumed that the kinds of models subjects initially construct are like that in Figure (5.4b). Only ab elements are represented in the model and consequently the predictions from the model would be that only ab elements are compatible with the rule. Experiment 5 showed that subjects gradually come to realize that -a-b elements are compatible with the rule. -a-b elements are not explicitly represented in the model and the model does not seem to distinguish between the status of -a-b elements which are compatible with a biconditional interpretation of the rule and -ab and a-b elements which are not, since none of these elements are explicitly represented in the model. Problems concerning the difference in status between a-b and -a-b elements also arise for the model in Figure (5.4a) as well as the model in Figure (5.4b).

It is useful at this point to consider how the difference in status between -a-b elements and a-b elements is established in other theories of reasoning. It has been mentioned that Piaget's theory would propose that not until the level of formal operations is a definite distinction made between -ab and -a-b elements which are compatible with an inclusion/implication relationship and a-b which are not, since only at this stage are the logical properties of implication understood. The elements comprising the combinations of formal operations represent negation explicitly so that an element -ab explicitly represents an entity which possesses attribute 'b' but does not possess attribute 'a'. This contrasts with the mental model representation in which it has to be inferred that the parenthetical 'b' element is not-a. In the Piagetian representation no distinction is made between the elements which make up a combination, for instance no distinction in status is made between the elements ab, -ab and -a-b which must be all "present" in order for an implication to be true. This is different from the distinction made in the mental model representation between the status of ab elements which are explicitly represented, -a-b elements which are not represented at all and -ab elements for which the attribute 'b' is explicitly represented but 'not-a' is only implied.

Theories of reasoning based upon Euler circles (Erickson, 1974) propose that in comprehending quantified statements subjects construct models of the statements which are isomorphic to the appropriate Euler circle representations of the statement. The Euler circle representation of "All A are B" would be represented by the two diagrams in Figure (5.5a) and Figure (5.5b). Figures (5.5a) and (5.5b) correspond (approximately) to the mental models in Figures (5.4a) and (5.4b) respectively. Elements outwith the perimeter of a circle in the

Euler diagram representation are members of the complement of the class represented by that cicle e.g. elements outwith the perimeter of the circle in Figure (5.5b) are members of not-A and not-B i.e. -a-b elements. The difference in status between -a-b elements and a-b and -ab elements in Figure (5.5b) is represented in terms of the existence of an area for -a-b elements but not for a-b or -ab elements.

The Euler circle model does not distinguish between conjunctive and biconditional interpretations of a universally quantified statement in terms of explicitly represented elements either: the distinction between a conjunctive interpretation and a biconditional interpretation in a Euler circle model could be captured by proposing that in a conjunctive interpretation the subject was not aware of the compatibility of elements outwith the circle in Figure (5.5b) with the truth of the statement while in a biconditional interpretation he was. The change from a conjunctive interpretation to a biconditional would presumably be captured by arguing that there is an increasing awareness of the logical properties of the representation. This increasing awareness of the logical properties of the representation could also account for different interpretations in the mental model theory although there does not seem to be a "zone" of the mental model representation which intuitively represents -p-q elements corresponding to the "zone" in the Euler diagram representation which represents -p-q elements. Mental models do not capture in such an intuitive way as Euler diagrams or Piagetian combinations the distinction in status between elements which are unacceptable (p-q) and elements with negative attributes (-p-q). It seems likely that this was what Johnson-Laird intended since he wanted to represent directly asserted information rather than implied information.

In mental model theory there is a difference between the way in which negative information is represented when it is explicitly asserted and when it is not explicitly asserted. Negative attributes which are not explicitly asserted are represented by omission: i.e. if an 'a' element for example is not identified with a 'b' element it is assumed to be 'not-b'. Thus in Figure (5.4a) the parenthetical 'b' element is inferred to be 'not-a'. Johnson-Laird presumably intended the theory to represent explicitly only asserted information from the sentence and that is why parenthetical 'b' elements are not explicitly defined as '-a' and why -a-b elements are not included in the model. Where a negative relationship is explicit, as in "No A are B", the negative relationship is represented by "fencing off" (Johnson-Laird's (1983, p.96) own term) the A elements from the B elements (Figure 5.6).

Venn diagrams, which are graphic representations of logical statements, similar to but not the same as Euler diagrams, allow the difference between a conjunctive interpretation and a biconditional interpretation of a universally quantified statement to be represented in an explicit way. Venn diagrams, unlike Euler diagrams or mental models, represent the universe of discourse of a problem by enclosing the problem space. A conjunctive interpretation of "All A are B" would correspond to Figure 5.7a. All elements in the universe of discourse would be ab elements. Gradually subjects would realize that -a-b elements were also possible discourse elements, Figure 5.7b. With time, the asymmetry of the universally quantified statement would be understood with the realization that -ab elements are also possible discourse elements, Figure 5.7c. This sequence of Venn diagram representations is very similar to the developmental sequence in the

Figure 5.6

a

a

a

---

b

b

b

Figure 5.7a

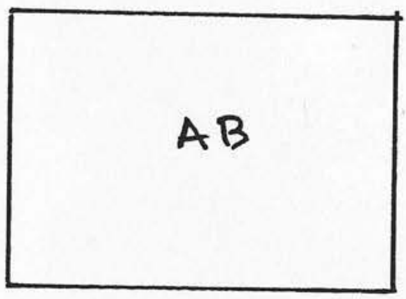


Figure 5.7b

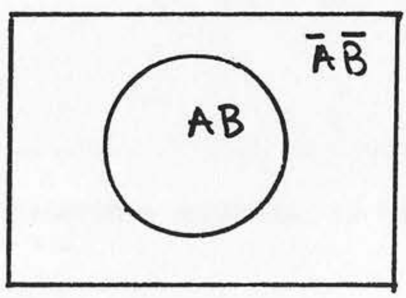
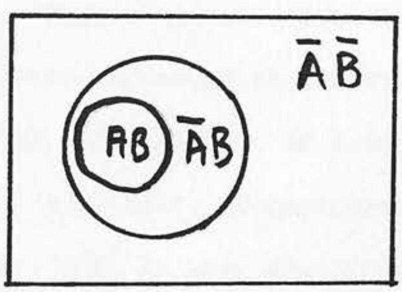


Figure 5.7c



ability to evaluate elements as possible elements of the universe of discourse as observed in the evaluation task. While the Venn diagrams represent different interpretations of the rule in the development sequence explicitly, the mental model theory (and the Euler diagram representations) would account for the different interpretations in the developmental sequence in terms of the subject's increasing awareness of the inferential properties of his representations.

In his principle of structural identity Johnson-Laird (1983) claims that :

"the structures of mental models are identical to the structures of the states of affairs whether perceived or conceived that the models represent."

Johnson-Laird argues that this is not true of other forms of meaning representation including truth tables which only capture truth functional relations between propositions, nor for Venn diagrams and Euler circles which have artificial structures created by logicians. Neither the standard type of propositional representations nor the type advocated by Johnson-Laird which is closer to the linguistic form of sentences have structures identical to the states of affairs they represent.

Despite Johnson-Laird's claim in the principle of structural identity, the inclusion of parenthetical elements in the mental model representations involves a degree of abstraction which would be absent from a representation of "states of affairs". The model in Figure (5.4a) is like a canonical representation of the universally quantified statement "All A are B". It is an abstraction from the structure common to the models depicted in Figures (5.4b) and

(5.8).The device of parenthetical elements introduces indeterminacy into the model. Johnson-Laird presumably introduced the parenthetical device into his mental model notation in order to obviate the problem of the combinatorial explosion which is a problem for theories of reasoning based on Euler diagrams. In combining the Euler diagram representations of two premises in a syllogistic reasoning task many different combinations of diagram would have to be considered in order to draw a valid conclusion: the number of different diagrams which would have to be considered would be at least the product of the number of diagrams required to represent the premises. Venn diagrams introduce indeterminacy into Euler diagram models in the same way as parenthetical elements introduce indeterminacy into mental models. In the Venn diagram a shaded area indicates that there are no members of that set while an area with no shading possibly has members. Figure 5.9 represents "All A are B". It is similar to an abstraction of Euler diagrams 5.5a and 5.5b in that it could represent either of these models and is thus similar to mental model 5.4a. The Venn diagram 5.9 is different from model 5.4a in that ab elements, a-b elements and -ab elements have a similar logical status since all three types of elements are possible elements while as mentioned previously in the mental model representation the status of all three elements is different: -ab elements are "asserted" to exist, -ab elements are possible and -a-b elements are either absent or inferrable, Johnson-Laird does not make clear which. The distinction in status between the different entities compatible with the truth of "All A are B" entailed by the mental model theory is psychologically compelling in that it is obviously much easier for subjects to agree that ab elements are compatible with the truth of "All A are B" than that -a-b or -ab elements are compatible with the statement.

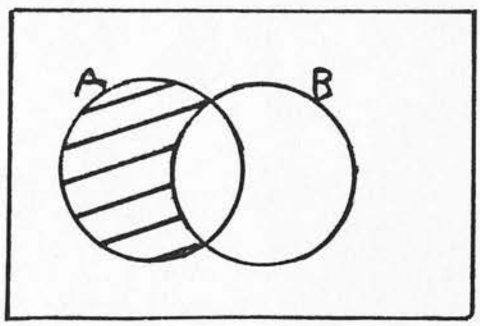
Figure 5.8

a = b

a = b

b

Figure 5.9



The abstraction implicit in the use of parenthetical items is not compatible with Johnson-Laird's claim that reasoning with mental models is "reasoning without logic". Johnson-Laird acknowledged that in drawing a conclusion in a syllogistic reasoning task an integrated model of the premises had to be constructed and tested. By systematic construction and exhaustive manipulation of the model and abstraction of a conclusion common to all constructed models, valid premises would emerge. It seems that in a similar way, the logical properties of a statement will "emerge" if the structure common to all possible models compatible with that statement is abstracted. By incorporating parenthetical items in the representations of statements it seems that Johnson-Laird is also incorporating the logical properties of these statements, i.e. properties which will determine valid inferences, into their representations.

THE EVALUATION AND SYLLOGISTIC REASONING TASKS

In the developmental literature on syllogistic reasoning it is frequently assumed that the 'matching' responses i.e. responses of valid to all four argument forms, Modus Ponens, Modus Tollens, Affirmation of the Consequent and Denial of the Antecedent, given by both children and adolescents are attributable to a biconditional interpretation of the major premise. However the results of the evaluation task show that primary school children apparently do not interpret conditional statements as biconditionals or class inclusion statements as class equivalence statements. Only 4% of response patterns by primary school children were consistent with such an interpretation. The response pattern analysis showed that where primary school subjects did adopt a consistent interpretation of the major premise it was a conjunctive/structure neutral interpretation: 28% of the consistent response patterns by primary school subjects were of this type. In addition primary school subjects were just as likely to accept p-q exemplars (29% accepted) as -pq (32% accepted) and -p-q (30% accepted) exemplars as compatible with the general rule showing that many subjects who did not adopt the conjúnctive interpretation still had difficulty with these exemplars.

There is no reason to suppose that children who interpret a conditional statement as a conjunction in the evaluation task should interpret it any differently in a syllogistic reasoning task. The problem with this suggestion is that most syllogistic reasoning studies have interpreted the response patterns of children of this age as reflecting a biconditional rather than a conjunctive interpretation. O'Brien and Shapiro (1968), for instance found that 6-8 year olds gave biconditional responses when reasoning with items like:

If this is Room 9 then it is 4th. grade. This is not Room 9. Is it 4th. grade? (a) YES (b) NO (c) NOT ENOUGH CLUES

Most young children responded "NO" on this (D.A.) argument form and responded according to the biconditional interpretation on the other argument forms too: "YES" to M.P. and A.C. and "NO" to M.T.

The close relationship between the conditional evaluation task and the conditional syllogism task has been discussed by Marcus and Rips (1979) who regard the evaluation task as "a kind of conditional syllogism in reverse". They argue that the evaluation task involves an inference from the components of the conditional to the conditional itself while in the syllogistic reasoning task an inference is made from the conditional to one of is components. When described in this way the difference between the evaluation task and the syllogism task seems similar to the difference between an inductive reasoning task and a deductive reasoning task. However, as has been discussed in the introduction to this chapter, the evaluation task is sometimes presented as a deductive reasoning task involving the evaluation of a specific instance with respect to a general rule and sometimes as an inductive reasoning or hypothesis testing task involving the evaluation (testing) of a general rule in the light of specific instances.

In view of the similarity between the evaluation and the syllogism task, Marcus and Rips expressed some surprise at the difference in the relative proportions of conditional and biconditional response patterns compared with inconsistent responses which adult subjects make in the evaluation task, 83% conditional responses, (Johnson-Laird and Tagart, 1969) and the syllogistic reasoning task in which Taplin and Staudenmayer (1973) found that only 42% of response patterns were conditional. In a within subjects design Marcus and Rips replicated the difference in proportions of consistent response patterns on the two tasks: 69.8% of responses on the evaluation task and only 32.1% on the syllogism task were conditional, with only 12.3% of responses on both tasks biconditional and 17.9% and 52.5% on the respective tasks logically contradictory. Marcus and Rips argued that the different results for the two tasks cannot be attributed to different interpretations of the conditional statement since the statements used in the two tasks were the same. Differences in response were attributed instead to the different inferential requirements of the two tasks: the evaluation task has less complicated inferential requirements than the syllogistic reasoning task.

While Marcus and Rips acknowledged that the evaluation task has inferential requirements of its own, Sternberg made a more radical claim that his "encoding" task, which is similar to an evaluation task, involves only linguistic encoding. Sternberg argued that the method of partial tasks, one of the methods of task decomposition, can be applied to the analysis of reasoning tasks to separate logical from linguistic factors in the development of reasoning abilities.

The rationale of the method of partial tasks as it applies to the analysis of reasoning problems is this: the whole task, i.e. the traditional syllogistic reasoning task, involving both linguistic encoding and logical inference, and the partial task, the encoding task which requires only linguistic encoding, are both presented to the subjects. If the analysis of errors on both tasks reveals that the whole task is not much more difficult than the partial task it can be concluded that the difficulty in reasoning is primarily one of linguistic encoding of the premises; if the whole task is much more difficult than the partial task it can be concluded that most of the difficulty in reasoning is associated with the logical combination of information.

Sternberg (1979) presented children from grade 2 (7.5 years) to college students (19 years) with either the encoding task or the combination task (corresponding to the evaluation and syllogism tasks). The premises included conjunctive, disjunctive, biconditional and conditional relationships as well as affirmation and negation of attributes. Overall subjects made more errors on the combination task than on the encoding task: 70% response patterns in the combination task and 59% of response patterns in the encoding task did not conform to the correct model pattern for that connective. Although a paired t-test on the overall means from the two tasks for each logical connective was significant, ($t(4)=2.33$, $p<0.05$), Sternberg concluded that most of the errors occur in the initial encoding of the logical connectives and consequently that the major difficulty in reasoning is the linguistic interpretation of the connective.

It is interesting to note that had Marcus and Rips analysed their evaluation and syllogistic reasoning tasks in terms of the method of task decomposition they would have probably concluded, as they apparently did anyway, that logical rather than linguistic factors are largely responsible for errors in syllogistic reasoning since 30.2% of their evaluation responses and 67.9% of their syllogistic reasoning responses did not conform to the correct model pattern; in other words a large percentage of the errors in the syllogistic reasoning task occur at the logical inference stage rather than the linguistic interpretation stage. Sternberg would presumably account for the difference between his results and those of Marcus and Rips by arguing that the difficulty for children is one of linguistic interpretation but once they can interpret the statement reasonably well the major difficulty becomes one of logical inference. Sternberg concedes that "it may be that some logical factors are involved in the encoding task" but he claims that these are of minor importance compared with the linguistic factors involved.

According to the logic of the method of partial tasks it would not be possible for the subject to respond in the whole task in a developmentally more advanced way than on the partial task. However, a comparison of the results of the evaluation task of experiment 5 and performance by subjects of the same age, between 6 and 11 years, on syllogistic reasoning problems indicates that children of this age are considered to perform in a syllogistic reasoning task according to a developmentally more advanced interpretation, i.e. the biconditional, than subjects of the same age in the evaluation task who predominantly adopt a conjunctive interpretation. This will be discussed in Chapter 6.

SUMMARY

In this chapter the conditional evaluation paradigm, another technique which purports to assess comprehension of the conditional, was discussed. The task requires the subject either to evaluate the truth of a conditional statement or rule in the light of different instances which have specific relationships to the rule (Johnson-Laird and Tagart, 1969) or to evaluate the compatibility of the instances with the rule (Marcus and Rips, 1979). A modified version of the task including both class and conditional statements (rules) was designed so that young children would be able to understand the task requirements. Prior to the age of 12-13 years subjects consistently accepted only instances with the attributes mentioned in the general rule as compatible with that rule even when the rule was expressed as a class statement. The young subjects who did not accept only pq exemplars were just as likely to accept p-q exemplars as compatible with the rule as -pq and -p-q exemplars. Significant changes in the patterns of exemplars accepted were found between primary and secondary school with subjects realising that -p-q elements were compatible with the rule. Secondary subjects still had difficulty in using the logical structure of the rule to work out which exemplars were compatible with the rule and the asymmetry of the rule particularly caused difficulty: although by third year secondary subjects were just as likely to accept -pq elements as -p-q with the class rule, -pq exemplars were still frequently judged as unacceptable with the conditional rule. The changes in the pattern of exemplars accepted as compatible with the rule at adolescence were regarded as consistent with Piaget's proposal that a qualitative change in the child's thinking occurs at adolescence with a "reversal in direction between reality and possibility". The primary subjects seem to regard the rules as descriptions of a determinate reality while the formal operational subjects can appreciate the hypothetical nature of the rule. The relevance of Johnson-Laird's mental model theory to the evaluation task was also considered.

CHAPTER 6

CONDITIONAL AND CATEGORICAL SYLLOGISTIC REASONING

It is well established within the literature on the acquisition of deductive reasoning abilities that both children and adolescents are much better at evaluating conclusions to valid argument forms, e.g. Modus Ponens (detachment), and Modus Tollens (particular contraposition), than to invalid argument forms, e.g. Affirmation of the Consequent (particular conversion), and Denial of the Antecedent (particular inversion). Ennis and Paulus (1965), for instance, presented their subjects (grades 5, 7, 9 and 11, mean ages 10, 12, 14 and 16 years) with all four argument forms and found a very low incidence of correct reasoning with A.C. and D.A.: only 2% of grade 5 and 3% of grade 11 subjects, and 3% of grade 5 and 12% of grade 11 subjects were correct on A.C. and D.A. argument forms compared with 51% and 62% for M.P. and 30% and 35% for M.T. Roberge (1970) found that his subjects performed better than this on Modus Ponens (grade 4: 53% correct; grade 10: 100% correct) and Modus Tollens (grade 4: 35% correct; grade 10: 65% correct) but that they were almost as poor as Ennis and Paulus' subjects on A.C. (grade 4: 2% correct; grade 10: 19% correct) and D.A. argument forms (grade 4: 2% correct; grade 10: 10% correct).

Although other studies (O'Brien and Shapiro, 1968; Kuhn, 1977) have shown better performance by young children on all four argument forms, the essential difference in performance between valid and invalid argument forms remains. What is disputed however is how this difference should be interpreted. O'Brien and Shapiro (1968), Hill (1960), and Suppes (1965), would agree with Ennis et al (1969) that the ability of children between 6 and 8 years to reason correctly with

Modus Ponens and Modus Tollens argument forms is "counterevidence to Piaget's claim that children are incapable of doing propositional logic" (p.70). Ennis (1975, 1976) argued that the experimental evidence which indicates that young children can draw correct conclusions with certain propositional logic argument forms indicates that it is better not to treat class and propositional logic, as Piaget does, as 'unitary wholes' which are attained at a certain stage of development. As was mentioned in the introduction to this thesis Ennis proposed instead an alternative approach to the study of logical reasoning in which he proposed that facility with different logical arguments was acquired at different stages of development.

Knifong (1974), however, did not accept the developmental distinction in performance on valid and invalid argument forms as incompatible with Piagetian theory. Rather, he claimed that Piaget's theory would have predicted such results and he gave an account of some of the experimental data based upon Piaget's theory. Knifong argues that the pattern of responses of young children on the 4 argument forms are exactly what Piaget's theory of transductive reasoning would have predicted. The distinguishing characteristic of the transductive reasoning of the young child is that it is reasoning based on the juxtaposition of elements which "go together". Although some authors (e.g. Taplin et al 1974) have indicated that transductive reasoning is initially based on a conjunctive interpretation of the relationship between elements, Knifong proposes that the child understands the relationship rather as a biconditional: when one element is present, the other is also present and when one element is absent, so is the other. The difference between this and the conjunctive response pattern is that under a conjunctive interpretation, p and q are always presumed to co-occur, but the absence of both p and q is not understood to be compatible with the truth of the conjunction. Thus, when presented with a conditional statement like:

"If this is room 9 then it is fourth grade" (Hill, 1960; O'Brien and Shapiro, 1968)

the assertion of a minor premise will lead to the acceptance of the conclusion, and the denial of a minor premise will lead to the denial of a conclusion: this is a biconditional response pattern.

Knifong regards transductive reasoning as reasoning based upon a "limited" combinatorial system. He proposes that instead of understanding a conditional 'If p then q' as equivalent to the propositional combination corresponding to implication (p.q) V (-p.q) V (-p.-q) the transductive reasoner understands the conditional as equivalent to the propositional combination (p.q) V (-p.-q), which is the propositional combination corresponding to the biconditional or equivalence relationship. The major problem with this proposal is that biconditional reasoning is also propositional deduction since the biconditional is also one of the propositional logical connectives and consequently, in Piaget's theory, would also be claimed to be a formal operational acquisition.

Bereiter, Hidi and Dimitroff (1979) argue that transductive reasoning and propositional biconditional reasoning generate the same response patterns to the 4 argument forms, but that these responses are generated by qualitatively different types of reasoning. They argue that the transductive reasoner does not understand the logical nature of logical connectives, but regards elements as "going together" (Flavell, 1963) simply because of their co-occurence in a statement. The young child makes associative "and-connections" (Flavell, 1963) between elements, rather than logical implication, or, for that matter, logical equivalence connections. Bereiter et al proposed that biconditional deductive, but not transductive, reasoners would be able to distinguish logically connected assertions from logically unconnected assertions. They devised a task in which subjects were presented with a conditional rule, followed by a question, and some clues to help them answer the question:

e.g. RULE: If it is a hot day then Judy will wear her blue skirt.

CLUES: (1) Blue is Judy's favorite colour.

(2) It was a sizzingly hot day.

(3) Judy looked lovely in her blue skirt that day.

QUESTION: Did Judy wear her blue skirt?

One of the clues was a critical clue (in this example, 2), from which the subject could make a deductively valid inference (if the deduction was not made, the following clue asserted the answer as in 3). The other clues were informationally relevant, but provided no deductively valid answer to the question. The presentation of the critical clue was varied across the 3 positions and the child was required to respond only when he was sure of the correct answer. The results showed that, as predicted, 7 year-olds were very poor compared with 11 year-olds at discriminating logically conclusive from logically irrelevant evidence with only 25% of 7 year-olds but 81% of 11 year-olds consistently responding to the critical clue. Bereiter et al also noted that 2nd. graders often drew conclusions from the rules alone, treating the rule as an affirmative rather than a conditional statement.

Taplin, Staudenmayer and Taddanio (1974), looked at the abilities of subjects from 9 years to 17 years to reason with abstract conditional sentences. They presented their subjects with problems like:

> If there is a Z then there is an H.
> There is a Z.
> There is an H.

The subject was required to say, given that the first two sentences were true, whether the 3rd was always, sometimes, or never true. Both affirmative and negative versions of the 4 different argument forms were used. Responses to particular argument forms which were statistically consistent across 12 replications were used to try to infer a "truth function" for the major premise. Taplin et al found that 3 different truth functions could be inferred. The percentage of subjects inferred to have a truth function did not change with grade, but the predominance of each truth function changed with grade: for grade 3 subjects (average age 9 years) the conjunctive and biconditional interpretations were equally prevalent, but by grade 5 (average age 11 years), the biconditional was predominant, with few response patterns being classified as conjunctive. The biconditional response was, in fact, predominant for subjects at grade 5, grade 7 (average age 13 year 7 months), grade 9, and grade 11 (average age 17 years), but by grade 9, some subjects began to respond according to the conditional, and the number of conditional response patterns increased slightly at grade 11. The predominance of the biconditional response pattern has been observed in many studies (Knifong, 1974; Bereiter et al, 1979). The conjunctive interpretation by the younger subjects had not been observed in previous studies, since most studies looked at correct response on particular argument forms, rather than considering responses as comprising a response pattern (O'Brien and Shapiro, 1968; Ennis et al, 1969).

Taplin et al found significant differences in the distribution of response patterns between grades 3 (9 years) and 5 (11 years), due to a decrease in conjunctive accompanied by an increase in the biconditional response pattern and between grades 7 and 9, due to a decrease in biconditional accompanied by the start of conditional interpretation. Taplin et al argued that these results could be explained in two different ways:

a) The results could be interpreted as supporting Piaget's theory, since reasoning becomes increasingly logical with increasing age, and the largest increases occur at the age ranges corresponding to Piaget's concrete operational and formal operational stages. It is interesting that, whereas most researchers regard changes in performance at the transition from one stage to another as compatible with Piaget's theory, Taplin et al regard changes in performance during the period of a particular stage as consistent with Piagetian stage theory. The increasing logicality explanation rests upon the assumption that the conjunctive, biconditional and conditional interpretations of the conditional are increasingly correct approximations to the meaning of the conditional.

b) Taplin et al also propose an alternative linguistic explanation of their results: reasoning at all ages is logical, but the meaning of the connective is different for subjects at different grades. The rationale for this explanation is that the number of subjects responding in a truth-functional way does not change across grade, but their consistent response is based upon different interpretations of the major premise. This position will be recognised as a rationalist position that inference is always carried out in accordance with logical principles, and errors in reasoning occur at the stage of interpretation of the premises.

Taplin et al go on to argue that some of Piaget's earlier work (Piaget

1926,1928) supports the interpretational change explanation rather than the increasing logicality explanation. In his earlier work Piaget looked at how children understand causal and implicative connectives such as "since" "because" etc. He found infrequent use of such connectives before 7-8 years: when they were used they were used to express co-occurence or juxtaposition of elements rather than implication or causality. Taplin et al argue that their results support Piaget's early contention that initially the implicative connective is understood as expressing a conjunctive connection between events or elements; the biconditional interpretation gradually appears as subjects realize that not only do events p and q co-occur but the negations of these events co-occur.

The difference between Taplin's and Piaget's (early) accounts of the change in meaning of the conditional with increasing age is that Taplin, but not Piaget, proposes that the subject is, at all ages, responding according to a truth-functional interpretation of the conditional, albeit an incorrect interpretation. Certainly the response patterns are consistent with truth functional interpretations but, as Bereiter et al have shown, the same response patterns can be produced by subjects who do not have an understanding of the logical necessity of the relationship but are merely responding in a transductive/intuitive way.

The increasing logicality explanation is not really an alternative explanation of the results in the way that Taplin et al present it; rather it is a description of the results and is compatible with both the linguistic explanation and Piagetian theory. The increasingly complex interpretations of the conditional are necessarily increasingly correct when assessed with respect to the normative criterion of propositional logic. It does not seem particularly useful to say, as Taplin does, that a subject consistently interpreting a conditional as a conjunction is just as logical as a subject consistently interpreting the conditional as a conditional since the former interpretation is defined as incorrect and the latter as correct. There are two factors which should be considered here - logical consistency and logical correctness. Taplin et al considered it important that the percentage of subjects classified as consistent did not change across grade; however just because a subject is consistent does not mean to say that he is correct.

In arguing that developmental changes in conditional reasoning performance are linguistic Taplin et al acknowledge that they do not distinguish between linguistic and cognitive development and that their results do not allow them to do so. Staudenmayer and Bourne (1977) contrast a Piagetian explanation of the developmental changes in interpreting connectives with linguistic development theories like that of Brown (1973). Piaget, of course, would attribute changes in the interpretation of logical connectives to changes in the kinds of cognitive operations available to the child: consistent logical interpretations of the propositional connectives would not be found until formal operations when the cognitive operations corresponding to the combinations of propositional logic are acquired. Brown, on the other hand, proposes that changes in the interpretation of the logical connectives depend upon the relative frequencies of occurrence of these connectives in the child's experience and also upon the semantic complexity of the particular relation. More complex relations like the conditional will be understood at a later age than simple relations like conjunction.

Staudenmayer and Bourne (1977) point out that these two theories attribute constraints in reasoning to a) the linguistic interpretation of the connectives (Brown) and b) the availabiliy of the cognitive operations to make use of this interpretation (Piaget). They argue that rather than becoming embroiled in the linguistic/cognitive primacy debate it is more instructive to ask: "What cognitive operations are necessary to make a particular interpretation?" (p. 617). While this does indeed seem to be an appropriate question it should be pointed out that Piaget did approach the problem in this way. For Piaget as for Staudenmayer and Bourne the cognitive operations available determine the linguistic interpretation and hence the permissible inferences.

Previous studies (Shapiro and O'Brien, 1970; Taplin et al, 1974) had indicated that children at certain ages have preferred interpretations of the conditional connective. Staudenmayer and Bourne asked whether there were any fundamental constraints in the cognitive operations available to the younger child or whether, as Brown suggests, experience with a particular interpretation of the connective might be all that is required to promote comprehension of that connective.

Staudenmayer and Bourne carried out an experiment similar to that of Taplin et al but in which the subjects (grade 3: ages 8.5 years, grade 6: 11.5 years and grade 9: 14.5 years) received feedback after every trial that was consistent with one of the three different interpretations of the conditional - conjunctive, biconditional and conditional. The results showed that, as expected, performance improved with increase in age : overall 31%, 58% and 73% of 3rd. 6th. and 9th. graders responded correctly according to the appropriate feedback. However while all subjects performed reasonably well with the conjunctive feedback (56.5%, 77.3% and 88.9% correct repectively) and grade 6 and grade 9 subjects performed well with biconditional feedback (31.0%, 85.7% and 96.4% correct) subjects at all grades performed poorly when feedback was in accordance with the correct conditional interpretation of the premises (5.3%, 11.1% and 33.3% correct).

Staudenmayer and Bourne argue that these results are not consistent with the linguistic experience interpretation since experience with a particular interpretation did not always lead to good performance with that interpretation. Of course this analysis rests upon the rather dubious assumption that feedback on twelve replications of a particular interpretation constitutes the sort of experience which would promote a facility with that interpretation. Staudenmayer and Bourne claimed that their results suggested that there is a fundamental limitation in the cognitive operations available to the younger child in his attempts to reason according to the feedback. Staudenmayer and Bourne explain the patterns of errors on the different forms made by subjects at different stages in terms of the information processing strategies that the children use in dealing with the problem and in doing so they essentially give an information processing interpretation of Piaget's theory.

Although syllogistic reasoning tasks are usually considered to be tasks involving formal operational thinking (Knifong, 1974; Staudenmayer and Bourne, 1977) it will be recalled from Chapter 3 that Kuhn (1977) regarded concrete operational thinking as the necessary and sufficient condition for correct response on a syllogistic reasoning task. Kuhn argued that the concrete operational child who can comprehend a class statement like "All P are Q" should also be able to understand a conditional statement like "If p then q" because of the logical equivalence of "If p then q" and "All P are Q.".

Kuhn suggested that the low incidence of correct reasoning found in most developmental studies of inference with conditional syllogisms might be attributable to the long and tedious nature of the tasks (Taplin et al, 1974, Roberge and Paulus, 1971) which may have caused subjects to lose interest and respond according to a simpler interpretation of the premises than they would adopt with more reflection. However Kuhn's replication of the Taplin et al procedure using only four items instead of the 96 used by Taplin et al produced similar levels of correct response.

Kuhn then suggested that a higher incidence of correct response might be obtained if the task were presented in a concrete conversational format rather than a formal written mode. Kuhn presented her subjects, grade 1 (6 years 7 months) to grade 4 (9 years 10 months) with all four argument forms with both class and conditional major premises. The relationship, between the terms p and q was contingent and plausible e.g.

All the people in Tundor are happy.
Jean does not live in Tundor.
Is Jean happy?

Kuhn did find a higher incidence of correct response on M.T. (80% correct), D.A. (72% correct) and A.C. (40% correct) argument forms which had been poorly answered in previous studies. Kuhn regarded her results as supporting her hypothesis that correct conditional reasoning is within the competence of concrete operational subjects and is facilitated by concrete conversational presentation. It was mentioned in the introduction to the thesis that Piaget acknowledged that concrete operational subjects can make correct propositional inferences provided that the propositions correspond to "sufficiently concrete representations". Presumably the content of the premises in Kuhn's task was of this nature. It is interesting to note that Kuhn found relatively high levels of correct response on the argument forms D.A. and A.C.. Rappoport (Ennis, 1976) argued that correct response on these argument forms requires consideration of two truth table contingencies and would consequently constitute a better test of formal reasoning than M.P. and M.T. which can be answered correctly by less sophisticated methods.

For Piaget it is the concrete content of the premises which enables the concrete operational subject to give correct responses on some propositional reasoning tasks. Although Kuhn also demonstrated that the concrete conversational mode of presentation of the task greatly improved performance on the syllogistic reasoning task, she also suggested that it was the concrete operational subject's comprehension of the logical properties and consequences of class inclusion which enabled him to generate correct inferences with the class statement and the logically equivalent conditional. However the fact that the improvement in performance occurred with the concrete/familiar content indicated that correct response was not mediated by an analysis of the formal properties of the class statement. The correct response to the problem above for example may be generated independently of any consideration of the major premise e.g. a subject's reasoning may go: "Jean does not live in Tundor. Jean may be happy or maybe not since some people are happy and some are not.'"

The results of the verification task and the evaluation task suggested that primary school subjects understood the quantification of class inclusion in an extensional situation only and that they could not work out from the linguistic statement alone the inferential properties of the statement. It seems more reasonable to suggest then that the ability of concrete operational subjects to respond correctly on a syllogistic reasoning task with class and conditional premises is due to the subject's ability as Piaget puts it, to "evoke the situation concretely" than to an understanding of the logic of inclusion: an understanding of the logic of inclusion however, presumably develops through abstraction of the logical properties common to all concrete problems.

In their "feedback" study Staudenmayer and Bourne gave their subjects feedback according to the conjunctive, biconditional and conditional * (See over) patterns shown in Table 6.1. In this study both affirmative and negative conclusions were presented and the different response patterns were clearly differentiated from each other. Traditionally the syllogistic reasoning task is presented in question form rather than as conclusions to be evaluated (Taplin et al, 1974; Staudenmayer and Bourne, 1977). It is interesting to consider what a conjunctive response pattern on the affirmative versions of the four argument forms only would be like. The response patterns for the biconditional and conditional interpretations are shown in Table 6.2. The response pattern for the conjunctive interpretation however is not so straightforward. Take the example of Taplin et al:

If there is a Z, then there is an H.

According to a propositional conjunctive interpretation this statement would be true only when there was a Z and there was a H. In all other

* Following the subject's response on each argument form he was informed orally about the correctness of that response according to one of the three response patterns.

TABLE 6.1

EIGHT FORMS OF THE CONDITIONAL ARGUMENT WITH RESPONSE CATEGORIES

ALWAYS(A), SOMETIMES(S) OR NEVER(N) CORRECT

ARGUMENT FORM	MINOR PREMISE	CONCLUSION	CONJ.	BICOND	COND
AFFIRMING	P	Q	A	A	A
ANTECEDENT	P	-Q	N	N	N
DENYING	-P	Q	N	N	S
ANTECEDENT	-P	-Q	N	A	S
AFFIRMING	Q	P	A	A	S
CONSEQUENT	Q	-P	N	N	S
DENYING	-Q	P	N	N	N
CONSEQUENT	-Q	-P	N	A	A

TABLE 6.2

FOUR FORMS OF THE CONDITIONAL ARGUMENT WITH RESPONSE CATEGORIES

YES(Y), NO(N) OR MAYBE(M)

ARGUMENT FORM	INTERPRETATIONS COND	BICOND	ALT. CONJ.
Modus Ponens If p q. p, q?	Y	Y	Y
Modus Tollens If p q. -q, p?	N	N	M
Denial of the Antecedent If p q. -p, q?	M	N	M
Affirmation of the Consequent If p q. q, p?	M	Y	Y

circumstances, if there was only a Z or only a H or if there was neither a Z nor an H, the statement would be false. The correct response to M.P. and to A.C. would be YES. Consider though the D.A. and M.T. argument forms: for both argument forms the information in the minor premise contradicts the information in the major premise. For subjects who understand the conditional as a conjunction, reasoning with these premises is like reasoning with the premises:

There is Z and there is an H.

There is not a Z.

Is there an H?

Strictly speaking subjects adopting a truth functional conjunctive interpretation of the conditional should refuse to respond arguing that the information as presented is logically contradictory. In practice it seems likely that the logical contradiction will not be appreciated particularly since it is the younger subjects who are presumed to adopt the conjunctive interpretation. The subject is trying to understand the problem as a whole and in doing so one of his presuppositions will be that a response is expected.

In many problems involving a conjunctive interpretation it is only relevant to the solution of the problem that the conjunction is true when both components of the proposition are true. In some inference problems though correct response based upon an understanding of the logic of the situation would also require an understanding of the conditions under which a conjunction is false.

Bereiter et al (1979) have pointed out that the younger subjects' biconditional response pattern is not based upon a truth-functional understanding of the problem but is generated by an "intuitive" or "transductive" analysis based upon the co-occurence of the elements in the proposition. Although this "transductive" interpretation is apparently like a conjunctive interpretation in that the proposition is only true when both components are true, it should not be identified with a truth-functional conjunctive interpretation since the transductive reasoner, unlike the truth-functional reasoner, does not consider the conditions under which the proposition is falsé to be relevant. The most likely responses to an "intuitive" or"transductive" interpretation would be responses in which the polarity of minor premise and conclusion were matched and this would produce a "biconditional" response pattern.

It was mentioned in Chapter 5 that the apparently more complex inferential requirements of the syllogistic reasoning task compared with the evaluation task made it unlikely that response on the syllogistic reasoning task would be developmentally more advanced than response on the evaluation task. However the predominance of the conjunctive response on the evaluation task and the evidence from syllogistic reasoning studies that the predominant response of younger subjects was biconditional seemed to contradict this view. It seems though that some subjects who would produce a conjunctive response pattern on the evaluation task would produce a biconditional pattern on the syllogistic reasoning task. Although the response patterns are the same as those generated by truth functional interpretations of the premises the response patterns are not necessarily generated from a truth functional interpretation of the premises but would in fact be very largely a function of the task structure and requirements.

When the major premise is a universally quantified proposition or a general conditional there is another possible response pattern which subjects might make to a conjunctive interpretation of the rule. Consider:

All the people in Tundor are happy.

A "conjunctive" interpretation of this would be like Bucci's "structure neutral" interpretation:

All the people live in Tundor and are happy.

As before M.P. and A.C. responses would be YES. As with the propositional conditional the information in the major and minor premises of the A.C. and M.T. argument forms would be contradictory for a conjunctive interpretation of the rule: The minor premise "Jean does not come from Tundor" contradicts the major premise which is understood as saying that everybody comes from Tundor and is happy. Given that the subjects presuppose that a response is expected they will probably understand the conjunction as domain specific. Only within the domain of the problem is it the case that all people live in Tundor and are happy. When a minor premise is negative the most likely interpretation will be that the negated element exists outwith the domain of the rule: Jean is not happy, therefore Jean must be outside the domain of the rule which specifies that everybody is happy and lives in Tundor. The most likely response to the negated elements is a matching one and consequently the conjunctive response pattern is indistinguishable from the biconditional. However a more sophisticated type of conjunctive interpretation might be found: since the negated elements are outside the domain of the rule the relationship between the attributes in the rule does not hold. If "Jean does not come from Tundor" she may or may not be happy. The response to D.A. and M.T. under such an interpretation would then be MAYBE rather than the matching NO response. It is interesting to note that Kuhn found a much higher rate of MAYBE response to D.A. argument forms (72% correct) than to A.C. (40% correct) and also a number of MAYBE responses to M.T. It is possible that the some of the correct MAYBE responses on D.A. and some of the MAYBE on M.T. were due to this type of conjunctive interpretation.

The evaluation task showed that there was no difference in the ability of primary school children to evaluate the compatibility of exemplars with a general rule whether this rule was a class statement or a conditional statement although at 3rd year secondary, around 15 years, subjects did find some exemplars easier to evaluate with respect to a class rule. Since performance on the evaluation task and the syllogistic reasoning task are related the results of the evaluation task suggest that it is unlikely that subjects at the stage of concrete operations will perform better on an inference task with class statements than with conditional statements, which some authors have suggested Piaget's theory entails (Ennis, 1975, 1976; Roberge and Paulus, 1971; Osherson, 1975).

Most studies which have compared the effect of class or conditional linguistic form on syllogistic reasoning performance (Hill, 1961; O'Brien and Shapiro, 1968; Roberge and Paulus, 1971; Osherson, 1974, 1975; Kuhn, 1977) have found no systematic effects of linguistic form. Hill found that 6 year olds found it easier to recognize valid conclusions (with M.P. and M.T. argument forms) with conditional than with class statements but by 8 years this difference had disappeared while O'Brien and Shapiro found no significant effect of linguistic form in their study in which they included all four argument forms.

Roberge and Paulus (1971) found a significant main effect of linguistic form with class reasoning significantly easier than conditional reasoning. However this difference was not consistent across all grades as the significant grade by linguistic form (type of reasoning) interaction showed: class reasoning was significantly easier for 6th. and 10th grade subjects but not for 4th. or 8th. grade subjects.

Osherson (1974) had subjects evaluate the validity of conclusions of fairly complex logical arguments formulated either in the language of classes or propositions e.g.

All the red jars and all the large jars have tacks.
Can you be sure that every jar that does not have tacks and is not large is not red?

The basic logical structure of all Osherson's problems was $(p \rightarrow q) \rightarrow (r \rightarrow s)$. Osherson found no differences in the abilities of either children or adults to evaluate the validity of conclusions for class or conditional logic problems.

Kuhn found no significant effect of linguistic form but a significant grade by linguistic form interaction in her study of syllogistic reasoning. Again however the results were unsystematic with first and third graders marginally better on conditional reasoning problems and second and fourth graders better on class reasoning problems.

The results of the evaluation task (experiment 5) showed that the predominant response pattern of children under 12 years was the conjunctive pattern. It is pertinent to ask how subjects who respond according to a conjunctive interpretation on the evaluation task would respond on a syllogistic reasoning task like that used by O'Brien and

Shapiro, Kuhn and others. It has been suggested that subjects who respond according to a 'conjunctive' interpretation in the evaluation task will probably respond according to a biconditional interpretation in the syllogistic reasoning task.

EXPERIMENT 6

THE SYLLOGISTIC REASONING TASK

METHOD

The general procedure and subjects were as described in Chapter 3.

The syllogistic reasoning task was similar to the syllogistic reasoning task used by Kuhn (1977) and by Shapiro and O'Brien (1970). The four different argument forms - Modus Ponens (M.P.), Modus Tollens (M.T.), Denial of the Antecedent (D.A.) and Affirmation of the Consequent (A.C.) were presented. The questions were always affirmative.

Half the subjects were presented with conditional premises and half with universally quantified premises. The problem was presented by means of a story. The story for the insect content was as follows; that for shape content is in Appendix E:

"One day when Mr. Jones was in his garden he noticed lots of insects. We do not have a picture of the insects but we do know that Mr. Jones said this about the insects in his garden:

(1) "In my garden all the big insects are black."

and we know that this sentence is true. We also know that there were three different kinds of insect in Mr. Jones garden. Mr. Jones picked up one of the insects from his garden and he said:

(2) "This insect is not black.

(3) Is this insect big?"

TABLE 6.3

THE FOUR CONDITIONAL SYLLOGISTIC ARGUMENT FORMS

MAJOR PREMISE: In Mr. Jones garden if an insect is big then it is black.

INSTANCE	ARGUMENT FORM	INTERPRETATIONS COND	BICOND	ALT. CONJ.
A) This insect is big. Is this insect black?	Modus Ponens If p q. p, q?	Y	Y	Y
B) This insect is not black. Is this insect big?	Modus Tollens If p q. -q, p?	N	N	M
C) This insect is not big. Is this insect black?	Denial of the Antecedent If p q. -p, q?	M	N	M
D) This insect is black. Is this insect big?	Affirmation of the Consequent If p q. q, p?	M	Y	Y

Given what you know about the insects in Mr. Jones garden do you think the right answer is YES, NO or MAYBE?

Remember Mr. Jones had three different kinds of insect in his garden and he said (1). He picked up an insect and said (2)(3)."

The problem was read aloud to the subject and the major premise (1) and minor premise and question (2) and (3) were presented on separate written cards. The four different argument forms were presented in random order on separate cards. The four different argument forms along with the correct responses for biconditional/conjunctive, conditional and "alternative conjunctive" interpretations are shown in Table 6.3.

RESULTS

Preliminary analysis revealed that neither of the main effects of content nor linguistic form nor any of the interactions of either factor were significant. In further analysis the data were collapsed across these factors.

Percent correct response to each argument type at each grade is shown in Table 6.4 and plotted in Figure 6.1. A 6 (grade) X 4 (argument form) analysis of variance with repeated measures on argument form was carried out on correct response (see Table 6.5). The main effects of grade $F(5,186)=17.62$, $p<0.001$, argument form, $F(3,558)=87.69$, $p<0.001$, and the grade by argument form interaction, $F(15,558)=5.15$, $p<0.001$, were all significant.

Tests of simple interactions revealed significant F ratios for adjacent grade comparisons only for the S1/S3 comparison,

TABLE 6.4

PERCENT CORRECT RESPONSE ON THE FOUR ARGUMENT FORMS ACROSS GRADE

CONDITIONAL

	MP	AC	DA	MT	TOTAL
P2	56	6	0	50	28
P4	75	6	6	72	41
P6	63	38	31	81	53
S1	100	31	44	94	69
S3	94	44	44	75	64
S5	94	63	88	56	75
TOT	80	31	35	72	55

CLASS

	MP	MT	DA	AC	TOTAL
P2	81	6	0	75	41
P4	88	13	13	63	44
P6	75	31	6	75	47
S1	97	50	75	91	58
S3	94	44	56	63	64
S5	100	63	75	56	73
TOT	88	29	30	70	54

TOTAL

	MP	AC	DA	MT	TOTAL
P2	69	6	0	63	34
P4	81	9	9	69	42
P6	69	34	19	78	50
S1	97	25	37	91	62
S3	94	44	50	69	64
S5	97	62	81	56	74
TOT	84	30	33	71	55

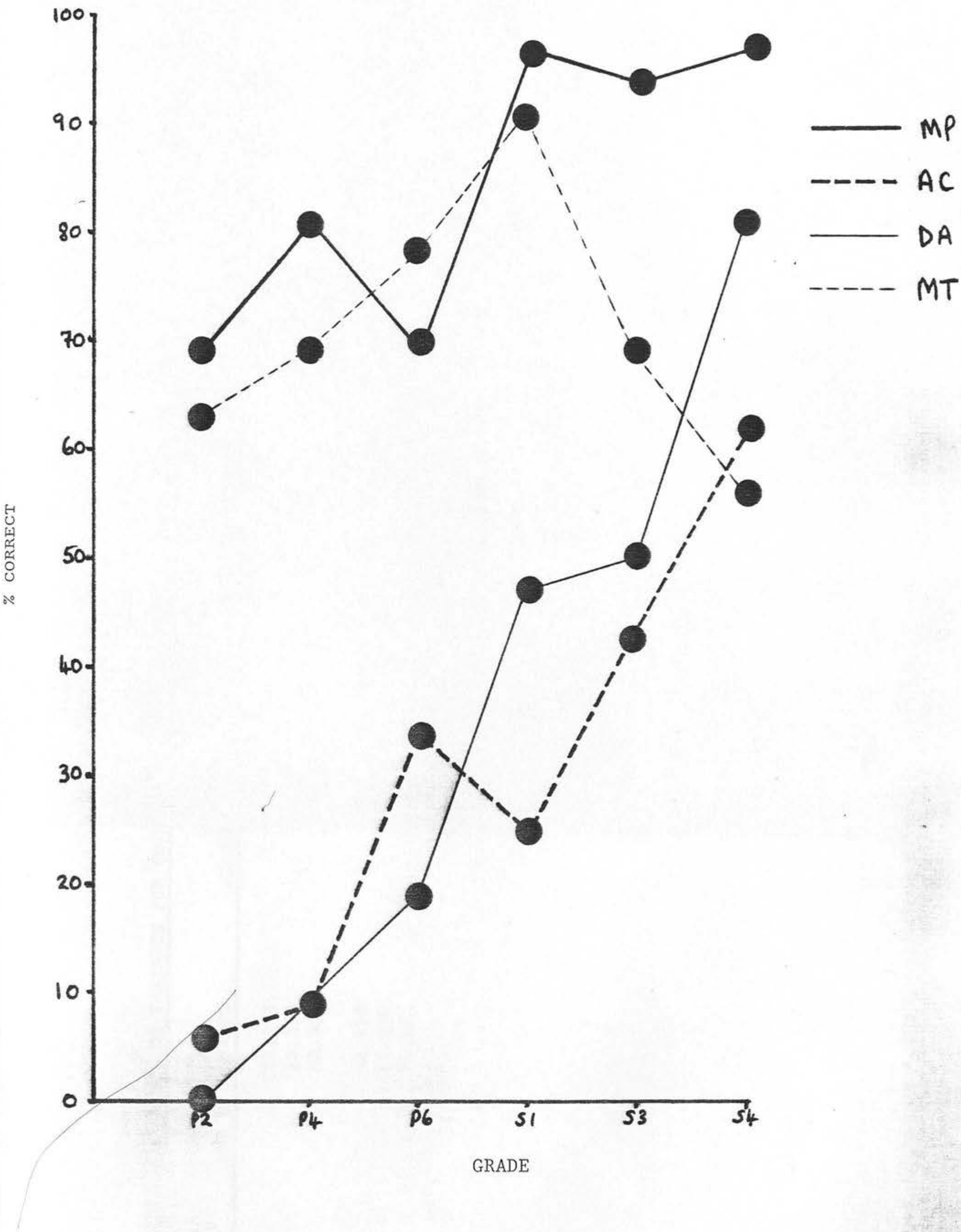


FIGURE 6:1

PERCENT CORRECT RESPONSE ON DIFFERENT ARGUMENT FORMS (DA,AC,MP,MT)

PLOTTED AGAINST GRADE

TABLE 6.5

ANALYSIS OF VARIANCE FOR CORRECT RESPONSE ON THE FOUR DIFFERENT ARGUMENT FORMS ACROSS GRADE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
MEAN	228.595	1	228.595		
GRADE	14.350	5	2.870	17.62	0.001
ERROR	30.305	186	0.160		
ARGFORM	42.618	3	14.206	87.49	0.001
AG	12.530	15	0.835	5.14	0.001
ERROR	90.602	558	0.162		

TABLE 6.5a

TESTS OF SIMPLE INTERACTIONS: F VALUES AND SIGNIFICANCE LEVELS FOR BETWEEN GRADE COMPARISONS

	P2	P4	P6	S1	S3	S5
P2	-	0.162 ns	0.352 ns	0.580 ns	3.438 p 0.025	35.208 p 0.001
P4		-	2.348 ns	0.354 ns	3.537 p 0.025	14.105 p 0.001
P6			-	2.508 ns	3.249 p 0.025	11.920 p 0.001
S1				-	3.247 p 0.025	13.115 p 0.001
S3					-	3.562 p 0.025

TABLE 6.5b

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF GRADE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
MP	2.875	5	0.575	3.55	0.01
AC	7.292	5	1.458	9.00	0.001
DA	14.359	5	2.872	17.73	0.001
MT	2.355	5	0.471	2.91	0.05
ERROR	120.907	744	0.162		

TABLE 6.5c

SUMMARY TABLE OF SIMPLE MAIN EFFECTS OF ARGUMENT FORM

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P2	12.625	3	4.208	25.97	0.001
P4	14.031	3	4.677	28.87	0.001
P6	7.563	3	2.521	15.56	0.001
S1	12.813	3	4.271	26.36	0.001
S3	4.844	3	1.615	9.97	0.001
S5	3.273	3	1.091	6.73	0.001
ERROR	90.602	558	0.162		

TABLE 6.5d

NEWMAN KEULS COMPARISONS BETWEEN ARGUMENT FORMS

ARG. FORM	A.C.	D.A.	M.T.	M.P.	r	s q(.95)(r,270)
ORDERED	0.302	0.328	0.708	0.844		s q(.99)(r,270)
MEANS						
A.C.	-	0.026	0.406	0.542	4	0.225
0.302		ns	**	**		0.275
D.A.		-	0.380	0.516	3	0.205
0.328			**	**		0.256
M.T.			-	0.136	2	0.171
0.708				*		0.226

$F(3,558)=3.25$, $p<0.025$, and the S3/S5 comparison, $F(3,558)=3.56$, $p<0.0001$, (see Table 6.5a). The results of tests of simple interactions on non-adjacent grades are also shown.

Tests of simple main effects of grade revealed significant F ratios for all argument forms (see table 6.5b). Tests of simple main effects of argument form were significant at all grades (see Table 6.5c).

Newman Keuls comparisons between means for argument forms showed that all comparisons were significant except that between D.A. and A.C. (Table 6.5d).

The results on the syllogism task were also analysed to discover whether there was any effect of the order of task presentation on response. As with the evaluation task it was thought possible that subjects who had previously answered the verification task would benefit from this experience and respond better on the syllogistic reasoning task than subjects who responded to the inference tasks first. However the absence of any order effects in the evaluation task indicated that presentation order would possibly not be important in the syllogistic reasoning task either. To establish whether there were any significant effects of presentation order was a 6 (grade) X 2 (linguistic form) X 2(order) analysis of variance was carried out on correct response. The results are shown in Table 6.6. Only the main effect of grade was significant, $F(5,168)=16.83$, $p<.001$. Response on the syllogistic reasoning task, like that on the evaluation task, was not affected by prior presentation of the verification tasks.

CLASSIFICATION OF RESPONSE PATTERNS

In addition to the analysis by argument form responses on the

TABLE 6.6

ANALYSIS OF VARIANCE FOR GRADE, LINGUISTIC FORM AND ORDER

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	184.62	191			
GRADE	57.401	5	11.480	16.83	0.001
LINGUISTIC FORM	0.005	1	0.005	0.01	ns
ORDER	1.172	1	1.172	1.72	ns
G X LF	3.777	5	0.755	1.11	ns
G X O	3.735	5	0.747	1.09	ns
LF X O	0.880	1	0.880	1.29	ns
G X LF X O	3.025	5	0.605	0.89	ns
ERROR	114.625	168	0.682		

syllogistic reasoning task were also analysed in terms of response patterns. Subjects were classified as adopting a biconditional/conjunctive, conditional or "alternative conjunction" response pattern if their responses accorded with the model patterns shown in Table 6.1. The distribution of response patterns found for subjects at each grade are shown in Table 6.7. Since the response pattern distributions for class and conditional premises were not significantly different, $\chi^2(3)=.894$, the data for class and conditional premises were collapsed together.

The predominant consistent response pattern for primary school children was the biconditional/conjunction pattern although for primary two subjects most responses were inconsistent. Although the comparison of response pattern distributions between primary two and primary four was not significant it did tend to significance, $\chi^2(3)=6.78$, $p<0.1$, reflecting a sizeable increase in biconditional and decrease in inconsistent response between P2 and P4. The only significant comparison between adjacent grades was between first and third year secondary, $\chi^2(3)=9.83$, p 0.025, and this was due to the decrease in biconditional accompanied by an increase again in inconsistent reponse. The S3/S5 comparison approached significance $\chi^2(3)=7.58$, p 0.1 and this reflected the change from inconsistent to conditional and "alternative conjunction" response patterns. The response pattern described as alternative conjunction was only found for 6th. year secondary school subjects.

DISCUSSION

The results of the syllogism task show that as in previous developmental studies of syllogistic reasoning ability with

TABLE 6.7

CLASSIFICATION OF RESPONSE PATTERNS

CLASS

	BICOND.	COND.	ALT. CONJ.	INCONSIST.
P2	4	-	-	12
P4	10	-	-	6
P6	7	1	-	8
S1	7	4	-	5
S3	4	3	1	8
S5	1	6	3	6
TOT	33	14	4	45

CONDITIONAL

	BICOND.	COND.	ALT. CONJ.	INCONSIST.
P2	5	-	-	11
P4	8	-	1	7
P6	9	-	-	7
S1	10	2	-	4
S3	2	3	-	11
S5	3	5	3	5
TOT	37	10	4	45

TOTAL

	BICOND.	COND.	ALT. CONJ.	INCONSIST.
P2	9	-	-	23
P4	18	-	1	13
P6	16	1	-	15
S1	17	6	-	9
S3	6	6	1	19
S5	4	11	6	11
TOT	70	24	8	90

categorical and conditional syllogisms primary school age children perform well on valid argument forms and poorly on invalid argument forms. The significance of the tests of simple main effects of grade for all argument forms reflected improvement in performance on M.P, D.A. and A.C. with age but for the M.T argument form correct response decreased following a peak at 3rd. year secondary. This decrease in correct response with Modus Tollens has been reported in previous studies (Kuhn, 1977; O'Brien and Overton, 1980). The improvement in performance with age on the invalid argument forms contrasted with the results of Ennis and Paulus (1965) and Roberge (1970) who found that older subjects also performed poorly on D.A. and A.C. If the non-significance of the simple main effect of argument form is taken as an index of mature performance not even the 5th. year secondary subjects had attained this.

The analysis of response patterns complemented the argument form analysis in showing a general increase in correct (implication) response with increasing age. The significant comparison between response pattern distributions for 1st. and 3rd. year secondary subjects which reflected a decrease in biconditional response paralleled the significant increases in correct response on the invalid argument forms at this age. The predominance of the inconsistent response pattern at 3rd. year secondary indicated that although subjects were beginning to realise that MAYBE was the correct response to some argument forms they still had problems in applying the MAYBE response correctly and overgeneralised its use to Modus Tollens. The decrease in correct response on M.T. contributed to the relatively low number of correct conditional response patterns even for 3rd year (19%) and 5th year (34%) secondary subjects.

An analysis of consistency of response across grade in experiment 6 showed a significant difference in the number of response patterns which can be classified according to one of the consistent response patterns across grade $\chi^2(5)=17.06$, $p<0.01$. This was not solely due to an increase in consistent response with increasing grade: rather there was an increase in consistent response up to S1 followed by an decrease in consistent response at S3 as subjects gradually realised that the biconditional response was incorrect. According to Taplin et al (1977) one interpretation of this kind of variation in consistent response across grade would be that it reflects a corresponding variation in "logicality" across grade. Rather than adopting this position it seems more reasonable to argue that there is an increase in logically correct response with increasing age but that subjects at certain ages adopt certain strategies which lead to consistent although logically incorrect response.

The predominance of the conjunctive response pattern as the consistent response pattern adopted by primary school subjects in the evaluation task might lead one to expect the same response pattern in the syllogistic reasoning task. It was mentioned in the introduction to this chapter that, strictly speaking, a subject who interpreted the major premise in the syllogistic reasoning task as a conjunction should regard the information in the minor premises for negative argument forms MT and DA as contradicting the major premise. For example a subject interpreting "In Mr. Jones' garden all the big insects are black." as meaning that in Mr. Jones' garden there are only big, black insects should regard the minor premise "This insect is not big" as contradicting the major premise since it was specified that the insect referred to in the minor premise came from Mr. Jones' garden. In fact no subjects claimed that the information presented in the MT and DA argument forms on the syllogistic reasoning task was contradictory.

The exemplar analysis in the evaluation task indicated that those primary school subjects who did not respond according to a conjunctive interpretation of the rule, accepting only pq exemplars, were just as likely to accept p-q (big, not black) exemplars as -pq (not big, black) and -p-q (not big, not black) exemplars as possible elements of the domain (possible insects in Mr. Jones' garden). Primary school subjects accepted on average 25% of -p-q exemplars, 31% of -pq exemplars and 27% of p-q exemplars as possible elements of the domain. Most subjects who were not classified as adopting a conjunctive response pattern on the evaluation task were classified as inconsistent. In contrast, on the syllogistic reasoning task primary school subjects gave either biconditional or contradictory responses. Knifong (1974), Ennis (1976) and Bereiter et al (1979) have pointed out that describing a response pattern as biconditional implies that the response pattern has been generated from an understanding (or rather a misunderstanding) of the logic of the problem. In the syllogism task, a logical biconditional understanding of the problem would require the subject to understand that only pq elements (big black insects) and -p-q elements (not big, not black insects) were compatible with the truth of the major premise (were possible insects in Mr. Jones garden) while -pq (not big, black) and p-q (big, not black) elements were not. Only if the subject understood this could he make a valid biconditional inference since only if he understood the logic of the situation could he draw a conclusion from the information in the premises which was necessarily true, for example he could infer that the insect referred to in the minor premise e.g. "This insect is not black" must be "not-big" since the only kind of insect in Mr.

Jones garden which is "not-black" is "not-big". However the evaluation task showed that primary school subjects did not understand the class or general conditional statement as equivalence/biconditionals. Subjects who were just as likely to accept p-q exemplars as -pq or -p-q exemplars as possible elements of the domain could not validly infer that an insect which was not-big must be not-black (-p-q) since it was just as likely to be black (p-q). It seems that rather than being generated from an analysis of the logic of the situation the biconditional response patterns of the primary school subjects were generated from an associative, "trasductive" or "matching" response like that described by Piaget (1967), Knifong (1974), Evans (1972). This response pattern is generated by matching the polarity of the attribute in the minor premise and conclusion. The shift in response from P2 to P4 from predominantly inconsistent to predominantly biconditional probably reflects an increase in the incidence of adoption of the matching strategy. It is not clear whether any biconditional responses were generated by an analysis of the logic of the situation rather than by a "matching" response. Since subjects on the evaluation task did not begin to differentiate between -pq, p-q and -p-q elements and evaluate them correctly until 1st. year secondary it seems unlikely that subjects' biconditional response patterns are logical biconditional before this.

Although the responses on the syllogistic reasoning task for primary school subjects did not seem to be predictable from responses on the evaluation task it was thought that the significant shift in consistent response from a predominantly conjunctive pattern at P6 to a predominantly biconditional pattern at S1 on the evaluation task reflected a fundamental change in the approach to the problem which would also lead to a change in strategy on the syllogistic reasoning task. The significant shift in response on the syllogistic reasoning task from 1st to 3rd. year secondary from predominantly biconditional to predominantly inconsistent response suggested that subjects at this stage realised that the matching strategy was inadequate and adopted a different strategy. It seems likely that subjects at this age began to realise that correct response depends upon an analysis of the logic of the situation but the predominance of the inconsistent response indicates that subjects found it difficult to work out exactly what the logic of the situation was. The decrease in correct response from S1 to S5 on the MT argument form also supports the view that subjects at this stage are beginning to appreciate the inadequacy of the matching strategy but the adoption of a more sophisticated response strategy taking cognisance of the formal properties of the argument would not be expected to lead to a decrement in performance. Although the older subjects are aware that the formal properties of the premises should be considered in making inferences the asymmetry of the inclusion/implication relation and negation still cause the subjects problems in actually making these inferences.

The distribution of response patterns at 5th year secondary indicates that there was still much uncertainty about correct response. Only at 5th year secondary was the response pattern described in the introduction as alternative conjunction adopted by a number of subjects. Subjects adopting this response pattern seem to interpret the major premise as meaning that everything within the prescribed domain is "a and b" but there are no constraints on the relationship between the attributes of elements outwith the domain of the major premise (outside Mr. Jones' garden). With this response pattern subjects respond YES to MP and AC is YES and MAYBE to MT and DA.

The absence of a significant effect of linguistic form in the evaluation and syllogistic reasoning tasks indicated that, as in the evaluation task, the ability to work out logical consequences of the inclusion relation in the absence of empirical or semantic support, like the ability to deal with the formal properties of implication, requires formal operational thinking.

Considering only the response patterns produced in a reasoning task rather than the abilities underlying the responses may lead to subjects who respond correctly on certain reasoning problems being inappropriately attributed with formal competence because their use of a less mature strategy leads to correct response (Bereiter et al, 1977; Staudenmayer and Bourne, 1979). There is, in other words, a danger of a false positive assessment of formal reasoning competence, as Wildman (1979) and O'Brien and Overton (1982) argue. The difficulties in distinguishing the reasoning abilities of the concrete and formal subject are described by O'Brien and Overton (1982) who point out the classification operations of the concrete operational subject can lead to correct response on some conditional reasoning problems and subjects at concrete operations may consequently be incorrectly credited with competence in reasoning.

The ability of young subjects to respond correctly on valid arguments in the syllogistic reasoning task is not regarded as evidence against Piaget's theory or, as Kuhn proposes, as evidence that concrete operations are the necessary and sufficient condition for syllogistic reasoning but, considered in conjunction with the responses on the evaluation task, is regarded as evidence of the use of a less sophisticated response strategy which leads to correct response only on some items. The significant changes in the exemplars accepted as compatible with the rule in the evaluation task in conjunction with the changes in response pattern on the syllogistic reasoning task at adolescence can also be accounted for in terms of different "means of reasoning" and are compatible with Piagetian theory.

It is relevant to recall Marcus and Rips's (1979) finding that their adult subjects produced a higher incidence of conditional response on the evaluation task (69.8%) than on the syllogistic reasoning task (32.1%). It seems that even when subjects can perform well on the evaluation task they still have difficulty in responding correctly to all four exemplars on the syllogistic reasoning task. Marcus and Rips attributed this to the greater difficulty of the syllogism task compared with the evaluation task. As in Marcus and Rips' study performance on the evaluation task was better than that on the syllogistic reasoning task with overall 68% of responses on the evaluation task correct and only 55% of responses on the syllogistic reasoning task correct. Following the method Sternberg (1978) used in assessing the relative difficulty of his encoding task (similar to the evaluation task) and his combination task (similar to the syllogistic reasoning task), correct response on the evaluation and syllogistic reasoning tasks was compared using a t-test for related measures on mean correct response collapsed across linguistic form for subjects at each grade. The significance of the t-test, $t(5)=7.39$, $p<0.0005$ on a one tailed test, showed that subjects performed significantly better on the evaluation task than on the syllogistic reasoning task.

The significant increases in reasoning ability on both the evaluation and syllogistic inference tasks at adolescence are compatible with the Piagetian view that there is a qualitative change in ability to reason with such problems around this age because of the acquisition of the combinatorial structure of propositional logic but the difference in levels of correct response on the evaluation and syllogistic reasoning tasks indicates that the inferential requirements of the tasks as well as their formal logical structure need to be considered in accounting for performance on the tasks and Piaget's model should be regarded as a model of logical competence rather than a model of performance.

It is interesting to consider how Johnson-Laird's mental model theory would account for the responses given in the syllogistic reasoning task. According to mental model theory, in comprehending the major premise "All A are B" the subject sets up a model like that in Figure 6.2a. He would then add on the information from the minor premise to this model. With the Modus Ponens argument form there is only one possible type of model which can be formed from the combined information from the major and minor premises (Figure 6.3a) and the conclusion which follows given "x is A" is "x is B". Subjects adopting the simpler representation of the major premise (Figure 6.2b) in which parenthetical "b" elements were not included would also respond correctly on Modus Ponens (Figure 6.3b).

There are two different types of model which are compatible with the A.C argument "x is B" (Figures 6.4a and 6.4b). The conclusion which follows from 6.4b is "x is A" and from 6.4a is "x is not-A". Consideration of both models compatible with the premises would lead to the correct response that no valid conclusion follows since x may be A or not-A. Johnson-Laird states a principle of mental model construction that in constructing mental models one is trying to establish as many identities as possible: according to this principle the model in Figure 6.4b would be established first. Subjects who failed to consider other models would be likely to conclude "x is A".

MENTAL MODELS OF THE SYLLOGISTIC ARGUMENT FORMS

Figure 6.2a

a = b

a = b

(b)

Figure 6.2b

a = b

a = b

Figure 6.3a

a = b

x = a = b

(b)

Figure 6.3b

a = b

x = a = b

Figure 6.4a

a = b

a = b

(b) = x

Figure 6.4b

a = b

a = b = x

(b)

Figure 6.5a

a = b

a = b

(b) = x

Figure 6.5b

a = b

a = b

(b)

x

Figure 6.6a

a = b

a = b

(b)

x

Figure 6.6b

a = b

a = b

x

Figure 6.6c

a = b

a = b

(a) = x

This was indeed the conclusion drawn by the majority of younger subjects. The results of the evaluation task suggest an alternative explanation of this conclusion. The results of the evaluation task indicated that it is more likely that subjects initially construct models of the major premise like that in Figure 6.2b than that in Figure 6.2a For younger subjects at least it seems likely that it is failure to construct an appropriate model of the premises rather than a failure to consider all alternative models which are true of the statement which leads to error with A.C. The correct inference, which was made by more subjects with increasing age, could only be generated from consideration of both models compatible with the truth of the statement.

Consider now the D.A. argument form "x is not-A.". Since the only elements represented in the model which are not-A are the parenthetical b elements it appears that the mental model theory would predict that the correct conclusion to D.A. is "x is B" (Figure 6.5a). This is clearly implausible since subjects rarely make errors of this kind. However Johnson-Laird builds into the model construction process the corollary of the principle that as many identities as possible should be established and that is that with negated elements as few identities as possible should be established. Thus the initial integrated model for D.A. would be as in Figure 6.5b and the conclusion would follow that "x is not-B". The simpler model of the major premise (Figure 6.2b) would also predict that this conclusion would be drawn. It is necessary to consider both models in Figures 6.7 and 6.8 in making the correct response.

Given the model construction principles stated above the Modus Tollens inference follows from the model in Figure 6.6a and the simpler model of the major premise, (Figure 6.6a) which the younger subjects construct would also predict the correct inference, "x is not-A". Consequently no change in response with age would be predicted for this argument form. However correct response on M.T. peaks at S1 and falls off again after this with an increase in the number of incorrect MAYBE responses. This decrease in correct response by the older subjects is difficult to explain in terms of mental model theory since the theory proposes that the model of the major premise is established first and the information from the minor premise added on. The MAYBE response on MT indicates that in making MT inferences subjects mistakenly suppose that there are parenthetical a elements which are not-b (Figure 6.6c) and they will consequently argue that if "x is not-b it may be a or not-a". Such errors would not be made by the younger subjects since they do not include parenthetical elements in their representations of the major premise but the errors suggest that subjects are still unsure about the constraints on the construction and manipulation of models. The inclusion of the parenthetical a elements would presumably only occur in contructing the integrated model of both premises since otherwise it would be incorporated into the models of other argument forms and this indicates that the interpretation of the major premise is influenced by the minor premise.

The mental model theory has not been explicitly addressed to developmental problems but can account for the relative ease of correct inference with valid argument forms and incorrect inference with invalid argument forms by younger subjects and the increase in correct response on invalid argument forms with increasing age in terms of the initial construction of oversimplistic models of the premises and the construction of more complex models with increasing age. The mental model account is similar to the Piagetian account in proposing that inferential ability is determined by the subject's ability to construct representations of the problem.

The mental model theory does not make any predictions about the age at which the increase in ability to deal with invalid argument forms would occur but there seems to be no reason why the evidence of qualitative changes in the ability to deal with the precise and flexible manipulation of language at adolescence should not be incorporated into mental model theory.

SUMMARY

In Chapter 6 categorical and conditional syllogistic reasoning studies were discussed. In experiment 6 subjects from 6 to 17 years were required to say whether conclusions to the four class or conditional syllogistic argument forms followed from the premises. As in previous studies no systematic differences in performance on categorical and conditional syllogisms were found and the poor performance of younger subjects on the invalid argument forms DA and AC contrasted with the good performance on the valid argument forms MP and MT. There were significant increases in correct response on both invalid argument forms and also on MP with increasing age but the indeterminate response was overgeneralised to the MT argument form by the older subjects.

The ability of young subjects to respond correctly on valid argument forms, the absence of a differentiation in performance on class and conditional logic problems, the relatively poor performance found on the invalid argument forms by older subjects and the decrease in correct response on MT have been found in other studies and have been regarded as evidence against Piaget's claims about the logical abilities of subjects at concrete and formal operations.

However it was argued that these results can also be given a Piagetian interpretation since Piaget proposes that the distinction between concrete and formal operational thinking should be drawn in terms of different means of reasoning. It was proposed that consideration of performance on different tasks could elucidate the different means of reasoning of subjects at different operational levels. The results of the evaluation task indicated that subjects understood that only

elements with the attributes mentioned in the rule were compatible with the rule. Given this interpretation of the rule on the evaluation task it seems likely that the biconditional response pattern of younger subjects in the syllogistic reasoning task was not logically determined, since a logical biconditional response would be generated from an understanding that -p.-q as well as pq elements are compatible with the major premise, but instead reflected a less mature "transductive" or "matching" response pattern. The increase in correct response on the evaluation task was accompanied by a realisation that the matching strategy was inadequate on the syllogistic reasoning task, and an increase in correct response on the invalid arguments brought about by a more effective strategy. The decrease in correct response on MT with increasing age also suggests a change in strategy.

CHAPTER 7

AN ATTEMPT TO INDUCE CORRECT SYLLOGISTIC REASONING

The significant improvements in reasoning ability on both experiments 5 and 6 at around 12 - 13 years, the stage identified by Piaget as the advent of formal operational thinking, were compatible with the general claims of Piaget's theory of the development of logical reasoning abilities. Certainly around adolescence individuals seem to have a more sophisticated understanding of the language of reasoning in that they seem to understand that many different states of affairs are compatible with a statement whereas younger subjects tend to regard statements as descriptions of a single reality. Adolescents seem to be able to cope much better than younger subjects with handling information which does not specify a unique state of affairs.

Although the performance of the 15 and 17 year olds on the syllogistic reasoning task was, in fact, superior to that found in many other studies they still made errors on 36% and 27% of items respectively. It was mentioned in Chapter 1 that poor performance by adolescents and adults on a variety of tasks which are regarded as formal operational has been understood by some (Neimark, 1975; Riegel, 1973; Wason, 1977; Wason and Johnson-Laird, 1972) to indicate that formal operational thinking is not achieved as universally as Piaget suggests. Studies which have compared performance on a variety of different formal operational tasks, including propositional reasoning tasks, have provided equivocal evidence regarding the interrelationships between performance on these tasks. Roberge and Flexer (1980) note that despite the observations in many studies of propositional logic abilities emerging during adolescence (Kuhn, 1977; Roberge, 1976;

Taplin et al, 1974; Roberge and Flexer, 1979, 1980; Roberge and Paulus, 1971; Staudenmayer and Bourne, 1977) analysis of interrelationships between performance on propositional logic tasks and other formal operations tasks have produced varying results (Bart, 1971; Kuhn, 1977; Roberge, 1976). If, as Piaget suggests, performance on formal operational tasks is mediated by an organized underlying structure there should be consistency in performance on these tasks.

Danner and Day (1977) have recently suggested that the poor performance of adolescents and adults on a variety of formal operational tasks (Blasi and Hoeffel, 1974; Papalia and Bielby, 1974; Tomlinson-Keasey, 1972) should not necessarily be seen as counterevidence to Piaget's proposals concerning the age of onset and universality of formal operations. They argued that frequently the requirements of the tasks used to assess formal operational thinking are ambiguous and because they are unclear about the aims of the tasks subjects do not perform on these tasks as well they could. Danner and Day argued that formal operational thinking could be regarded as "latent" and could be elicited simply by clarifying the task demands. Danner and Day and Stone and Day (1978) showed that by introducing a series of simple prompts performance by adolescents on several formal tasks could be significantly improved.

O'Brien and Overton (1980) have shown that "latent" formal thinking can also be demonstrated in conditional reasoning tasks but only for young adults (mean age 21 years 5 months) who should be well into the formal operations stage and not for adolescents (mean age 13 years 5 months) who are presumably just entering the formal operations stage. This is of interest because previous attempts to elicit successful conditional reasoning had generally been unsuccessful (Lunzer, Harrison and Davey, 1972; Staudenmayer and Bourne, 1979; Wason, 1968). Attempts to induce successful solutions to reasoning problems with abstract content by having subjects solve problems with the same logical structure but concrete content had also been unsuccessful since the ability to solve the concrete task did not transfer to abstract tasks (Johnson-Laird et al, 1972; Wason and Johnson-Laird 1972).

O'Brien and Overton used a method of "contradiction training", which had previously been found by Wason (1964) to be relatively effective in making the subjects aware of their self-contradictions, to improve conditional reasoning performance. Wason found that in certain circumstances subjects refrain from making incorrect inferences following the presentation of information which contradicts an incorrect inference previously made. In Wason's "contradiction training" paradigm subjects were presented with an incomplete conditional rule relating ages and salaries of employees of a hypothetical firm, e.g. "Any employee aged - years or more will receive a salary of at least $1900 a year." The subject was then presented with ten trials each consisting of a hypothetical employee of a particular age receiving a certain salary (see Table 7.1). The subject's task was to say for each example whether an inference could be made about the missing age in the rule. For exemplar (1), for example, it can be inferred that the age in the rule is greater than 26 years. This is because the salary in the exemplar is less than $1900. In order for the conditional rule to be true the age in the rule must be greater than the age in the exemplar, i.e. 26 years. Many subjects infer from exemplar (2) that the age in the rule must be less than 36 years: they reason that since the salary in the exemplar is more than that in the rule, the age in the rule must be less than that

TABLE 7.1

INFERENCES IN THE CONTRADICTION TASK ABOUT UNKNOWN AGE

TRIALS (EMPLOYEES)	SALARY($)	CONTRADICTION GROUP	CONTROL GROUP
		AGE AND INFERENCE	
1	1300	$>$26(V)	$>$26(V)
2	2400	$\ngtr$36(F)	$\ngtr$36(F)
3	1400	$>$27(V)	$>$27(V)
4	2300	$\ngtr$35(F)	$\ngtr$35(F)
5	1500	$>$37(V)	$>$28(V)
6	2200	$\ngtr$38(F)	$\ngtr$34(F)
7	1600	$>$40(V)	$>$29(V)
8	2100	$\ngtr$41(F)	$\ngtr$33(F)
9	1700	$>$43(V)	$>$30(V)
10	2000	$\ngtr$45(F)	$\ngtr$32(F)

Column 1 shows the salary earned by an employee at each trial. Columns 2 and 3 show the age of the employee, combined with the inference about unknown age.

V: Valid inference
F: Fallacious inference

in the exemplar. This inference is incorrect however since both pq exemplars (employees aged - years or more and earning at least $1900) and -pq (not aged - years or more and earning at least $1900) types of exemplar are compatible with the truth of the rule. When the salary in the exemplar makes the consequent of the conditional true no valid inference follows about the age in the antecedent of the rule.

In Wason's task adult subjects were given a sequence of trials and asked what could be said about the age in the rule. For subjects who made the invalid inferences on trials (2) and (4) that the age in the rule was less than 36 years and 35 years respectively, an inconsistency was introduced into the task on trial 5 where the valid inference could be made that the age in the rule would be greater than that in the exemplar. Wason found that when this "contradiction" (it is a contradiction only in sense that it contradicts an invalid inference that the subject has made) was introduced subjects were significantly better at withholding subsequent invalid inferences compared with a group who had not had the contradiction training. Wason argued that

"knowledge of fallaciousness lies within the individual's logical competence and the function of contradiction is merely to act as a prompt or cue." (Wason and Johnson-Laird 1977, p. 119).

The contradiction training seems to act as a hint to subject about the level of analysis which is appropriate for the problem: it alerts the subject to the fact that it is the formal properties of the rule which are important.

O'Brien and Overton (1980) used the "contradiction training" task to assess developmental differences in the effectiveness of contradiction. They were also interested in whether the effect of contradiction training generalised to performance on other conditional reasoning tasks - the selection task and the evaluation task. They found that the contradiction training was effective for their college student group (mean age 21 year 5 months) but not for their younger subjects (mean ages 9 years 9 months and 13 years 5 months).

Given the ineffectiveness of the contradiction training for the younger subjects it would have been surprising if these groups had shown any improvement on the selection and evaluation tasks and indeed contradiction training led to a significant improvement in performance on both the selection and evaluation tasks only for the college subjects. O'Brien and Overton interpret their results as indicating that young adults have the competence to understand the conditional relationship and that the appropriate formal operational strategies for the solution of particular tasks involving the conditional relationship can be invoked by the presentation of "prompts" in a similar way to that described by Danner and Day. In this case the "prompt" consisted of information which contradicted previous invalid inferences and this seemed to alert subjects to the asymmetry of the conditional and specifically to the invalidity of inferring p from q (Affirming the Consequent).

O'Brien and Overton (1982) extended their study to consider the transfer of the effectiveness of contradiction training to performance on the syllogistic reasoning task. The results of their 1980 study showing the effectiveness of contradiction training and transfer to performance on the evaluation task only for the older subjects (18 years 2 months) was replicated. Although O'Brien and Overton claimed that the effectiveness of contradiction generalised to the syllogistic reasoning task their results are not as clear-cut for this task. Performance of the contradiction training group on Modus Ponens and

Denial of the Antecedent was not significantly different from that of the control group at any grade. Only at 12th. grade was performance of the contradiction group significantly better than the control group on both Affirmation of the Consequent and Modus Tollens. The contradiction training was effective in improving performance specifically on the A.C. and M.T. argument forms, the argument forms which involve reversal of the order of the constituents of the conditional.

Using a different procedure from O'Brien and Overton, Falmagne (1980) found that subjects from 2nd. through 5th. grade (presumably about 7 to 12 years) could benefit from training and learn to make propositional inferences, in this case Modus Tollens inferences. Falmagne et al (1977) used a concept-learning paradigm in which their subjects were presented with a number of instances of a Modus Tollens inference e.g.

"If it is Tuesday, then Mary has gym today. Mary doesn't have gym today. Is it Tuesday? "

followed by feedback about the correct response. Two days later the subject was given more problems of the same kind in order to establish whether he had abstracted the logical form of the argument. Falmagne et al found that following training, subjects who had received training on the Modus Tollens inference performed significantly better than a control group. In addition the improvement in performance by the "training" group was greatest for nonsense problems such as "If Paul fibbles then he thabbles" which provided no clues from the content about inferential properties of the sentence.

The results of training studies are rather difficult to interpret since it is not always clear exactly what abilities the training tasks train. The training in Falmagne's task is more explicit, is of more obvious relevance to and is more specific to the acquisition of the Modus Tollens inference than the contradiction training in O'Brien and Overton's study is to performance on the syllogistic reasoning task. What is of interest in Falmagne's study is that the younger subjects can learn to abstract the logical form of the inference at all. It seems that the younger subjects can benefit from training only when the training is directly relevant to the ability being tested and sometimes not even then, as the ineffectiveness of the contradiction training for the younger subjects in O'Brien and Overton's (1980, 1982) studies showed.

Although Falmagne claimed to have shown that subjects who had not yet attained the formal operational stage could be trained to abstract the logical form of a Modus Tollens inference her results could be interpreted in another way. Falmagne included both positive and negative forms of Modus Tollens and presumably the correct response to the MT argument form was always either YES or NO. It was argued in Chapter 6 that correct response on Modus Tollens does not necessarily require an understanding of the formal properties of the inference and can be generated by a strategy of matching the polarity of the minor premise and the conclusion. This strategy would also be effective in generating correct response on Falmagne's task since the matching response is always correct on this argument form. Since Falmagne included only the Modus Tollens inference it seems probable that the feedback about correct response would reinforce the use of this successful strategy and consequently lead to higher levels of correct response than the control group who received no such feedback. Just as Rappoport (1967) and Staudenmayer and Bourne (1979) argued that the invalid argument forms constituted critical tests of formal operational thinking, so too in a training task it seems that a more convincing demonstration of the training of formal reasoning ability would be the training of correct response on invalid argument forms. Training like Falmagne's in which the subject is given feedback concerning the correct response would be likely to lead to consistent adoption of a correct MAYBE response: a demonstration of the training of a formal understanding of the problem would require that the subject could respond correctly to both valid and invalid argument forms.

The success of training obviously depends upon the complexity of the ability being trained. In Falmagne's task for instance it was argued that subjects at the stage of concrete operations had the ability to respond correctly on Modus Tollens anyway although not necessarily on the basis of a formal understanding of this inference. The lack of success of many studies attempting to induce correct reasoning with adults in the selection task for instance indicates that some abilities are difficult to train even in adults who are presumably well into the formal operational stage. Although adults also have difficulty with syllogistic reasoning, O'Brien and Overton showed that performance on specific argument forms on a syllogistic reasoning task could be improved by contradiction training.

Experiment 6 showed that one of the main problems which children and adolescents have with syllogistic reasoning problems is in responding correctly to the invalid argument forms. According to the competence model subjects at the stage of formal operations have the competence to understand the conditional but may fail to demonstrate optimum performance in a reasoning task for a variety of reasons. In the light of Piaget's claims about the stages of development of logical thinking it was thought that subjects at the stage of formal operations could be induced to solve syllogistic reasoning problems, and particularly invalid arguments, by having the subjects solve a concretized version of the problems in which they were presented with a pictorial representation of the major premise. It was predicted that formal operational subjects should have the competence to appreciate the significance of the concrete presentation and the concrete task would act as a prompt to successful solution on the abstract task. Since subjects at the stage of concrete operations are presumed to be able to solve reasoning problems when the problem is tied to concrete objects and relations it was argued that concrete subjects would be able to solve syllogistic reasoning problems when presented with a concrete representation of the premises but that concrete subjects would not have the underlying competence to appreciate the significance of the concrete task for solution of the abstract problems in the standard syllogistic reasoning task.

The verification studies had shown that, although primary school subjects make errors in verifying conditional statements, even the youngest subjects performed well in verifying true conditional statements (88% correct overall in experiment 2 and 73% correct in experiment 4). In an informal observation primary children were presented with syllogistic reasoning problems to solve with the aid of a pictorial representation of the major premise. The children apparently found the problems easy to solve. It was decided to use this "empirical syllogistic reasoning task" as a means of trying to teach children to respond correctly on the syllogistic reasoning task without actually giving them explicit feedback about the correctness of their responses as Falmagne had done.

EXPERIMENT 7

THE TRAINING TASK

METHOD

Materials

Four different eight-page booklets were prepared. Each booklet was to be answered in four "sessions" of two pages each with one problem per page. The four content areas for the problems were CLOCKS, TINS, FISH and BOOKS. The problems for the first and second sessions were identical except for the fact that session two was the "learning session" in which subjects answered the questions with the aid of a pictorial representation of the problem content which was consistent with the truth of the general rule. Although for any subject the content areas for session 3 problems were similar to those of sessions 1 and 2 the attributes of the elements in the rules were changed on session 3 so that the subject had to work out the correct answer and could not simply remember or copy the correct answers from sessions 1 and 2; for instance, if the general rule in sessions 1 and 2 was: "In Fred's shop if a clock has a round face it is at six o'clock." the rule for session 3 would be changed to: " In Fred's shop if a clock has a square face it is at twelve o'clock.". The content areas for session 4 problems were different to those of sessions 1, 2 and 3 so that, for example, if the two problems in sessions 1, 2 and 3 concerned FISH and BOOKS the two problems in session 4 concerned CLOCKS and TINS. The change of content on session 4 was to see whether any learning which occurred during session 2 generalised to problems with a different content.

On each page of the booklet there was one problem with eight examples. At the top of each page there was a description of the problem content followed by a general rule which the subject was told was always true for that problem. The general rule was expressed either as a general conditional or as a universally quantified statement using the quantifier "all". In each of the four sessions there was one problem of

Following the general rule there were eight different examples which were affirmative and negative versions of the M.P., M.T., D.A. and A.C. argument forms. The examples for the general rule: "In John's fishtank if a fish is green it has spots." are shown in Table 7.2. The Modus Ponens argument form with affirmative conclusion would be:

(A) Fish A is green.

Is this true? Fish A has spots.

(A) (1)YES (2)NO (3)MAYBE

The subject was required to evaluate whether the conclusion, which was underlined, was true given that both the general rule and the minor premises were true. Possible responses were YES, meaning that the conclusion must be true, NO, meaning that the conclusion could not be true and MAYBE, meaning that the conclusion might be true or might be false, there was not enough information to say. The eight instances were presented in different random orders for different problems except that the instances for the problems in sessions 1 and 2 were presented in the same order.

Denial used in negating the minor premise and conclusion was explicit negation rather than implicit or lexical negation, i.e. a particular element was described as either having or not having a specific

Table 7.2

Example of the problems used in Experiment 7
showing the eight different argument forms

This problem is about the fish in John's fishtank. In John's fishtank there are GREEN fish and YELLOW fish. Some of the fish have SPOTS on them and some have STRIPES. The rule for this problem is:

In John's fishtank, if a fish is green it has spots.
Remember this rule is always true.
Given that you know the rule and you know that:

(A) Fish A is not green.
Is this true? Fish A has spots.
(A)(1)YES (2)NO (3)MAYBE

(B) Fish B is green.
Is this true? Fish B has spots.
(A)(1)YES (2)NO (3)MAYBE

(C) Fish C does not have spots.
Is this true? Fish C is green.
(A)(1)YES (2)NO (3)MAYBE

(D) Fish D has spots.
Is this true? Fish D is not green.
(A)(1)YES (2)NO (3)MAYBE

(E) Fish E is not green.
Is this true? Fish E does not have spots.
(A)(1)YES (2)NO (3)MAYBE

(F) Fish F is green.
Is this true? Fish F does not have spots.
(A)(1)YES (2)NO (3)MAYBE

(G) Fish G does not have spots.
Is this true? Fish G is not green.
(A)(1)YES (2)NO (3)MAYBE

(H) Fish H has spots.
Is this true? Fish H is green.
(A)(1)YES (2)NO (3)MAYBE

attribute.

In previous syllogistic reasoning studies the individual in the minor premise and conclusion was referred to either by naming the individual e.g. "Jean lives in Tundor." or by using the determiner "this" to refer to the individual as in "This cat" or "This shape". It had been noticed in the previous study that some of the younger children were confused about whether "This shape" in the minor premise was the same as "This shape" in the conclusion. Obviously unless identity of reference is established between the individual in the minor premise and the conclusion valid deductions cannot be made. Using names is a good way of establishing identity of reference but was not very useful in this study since the learning session involved a pictorial representation of an inclusion/implication relationship and this would have to be relatively easy to depict. In addition there is a rather limited range of content for problems concerning attributes of people (or animals or other objects with names) which can be represented pictorially. In this study the device of referring to an individual as "Fish A", "Fish B" etc. was used to make it clear that the individual in the minor premise and conclusion of any argument was the same individual and that individuals in different arguments were probably, though not necessarily, different. While acknowledging that this departs from normal linguistic usage it was felt that subjects would be familiar with the use of variables from problems in maths.

Since subjects were tested in groups of five, four different booklets were prepared. These differed in whether the first problem in each session involved a class or a conditional rule and also in the content of the problems in the different sessions. For half the subjects the content of the problems in sessions 1 2 and 3 was CLOCKS and TINS and for the other half the content was BOOKS and FISH. Subjects receiving CLOCKS and TINS in sessions 1, 2 and 3 received BOOKS and FISH in session 4 and vice versa.

Examples of problems from the different content areas are shown in Appendix F. Also shown are the corresponding pictures of these content areas which were used in the empirical reasoning task of session 2. The pictures can be seen to be consistent with the truth of the relevant general rule.

Design

In many training studies the effectiveness of training is measured by comparing the performance of the experimental group with that of a control group who do not receive the intervening training (O'Brien and Overton, 1980, 1982; Falmagne, 1980; Barratt, 1975). Another way of assessing the effectiveness of training is to have the subjects act as their own controls and compare performance after training with that before training. In a study designed to elicit formal operational thinking, Danner and Day (1977) examined the effectiveness of prompts by comparing performance on a task before prompting with performance on another related task after prompting. This method of assessing the effectiveness of training, by comparing performance before and after training, was used in experiment 7.

In experiment 7 performance on session 1, prior to any training, was examined, since this was interesting in its own right as an index of performance on standard class and conditional syllogisms. A 5 (grade) x 2 (linguistic form) x 8 (argument form) analysis of variance with repeated measures on linguistic form and argument form was carried out on correct responses to the different argument forms on session 1. Performance on session 1 was also compared with performance on previous syllogistic reasoning studies.

Performance on the empirical reasoning task of session 2 was also of interest, since training was not likely to be effective unless performance on session 2 was significantly better than that on session 1. In order to compare performance on sessions 1 and 2, a 5 (grade) x 2 (session) x 8 (argument form) analysis of variance with repeated measures on session and argument form was carried out on correct responses to the different argument forms on sessions 1 and 2.

The major comparisons of interest in experiment 7 were those between session 1, which served as a measure of baseline performance, and the post-training sessions, session 3 and 4. A 5 (grade) x 2 (session) x 8 (argument form) analysis of variance with repeated measures on session and argument form was carried out on correct responses to the different argument forms on sessions 1 and 3 in order to establish whether there was any improvement in response following training for problems with a similar content to the original problems and, if so, whether it differed for the different argument forms and for subjects at different grades. A similar analysis of session 4 responses was carried out in order to show whether any improvement following training generalised to problems from different content areas.

Differences in performance on the different sessions in experiment 7 were also examined in terms of changes in response patterns.

<u>Subjects and Procedure</u>

Twenty children from primary five and primary seven of an Edinburgh primary school and twenty children from the second, fourth and sixth years of an Edinburgh secondary school took part in the experiment. Their average ages and age ranges were as follows 9 years 4 months (8 years 8 months to 9 years 9 months), 11 years 3 months (10 years 9 months to 11 years 10 months), 13 years 5 months (12 years 11 months to 14 years 0 months), 15 years 7 months (15 years 1 month to 16 years 2 months), 17 years 6 months (17 years 1 month to 18 years 1 month). The primary school was a feeder primary for the secondary.

Subjects were tested in groups of five and all subjects in any group received the same test booklet. The experimenter read the instructions aloud referring to the first page of the booklet to explain the layout of the problems. The general instructions were as follows (the problem referred to in the instructions is that shown in Table 7.2):

"I am interested in how children of different ages think and today I am going to give you some problems which require you to think very carefully but which are quite good fun to do.

In the booklet in front of you there are eight pages and each page has a problem on it with different examples. At the top of each page you are told what the problem is about, for instance, the problem on the first page is about the fish in John's fishtank. You are told that in John's fishtank there are GREEN fish and YELLOW fish. Some of the fish have SPOTS on them and some have STRIPES.

You are then given a general rule for the problem, which is underlined. The general rule is true for the whole problem, that is for all the examples on that page. In this problem the general rule is "In John's fishtank, if a fish is green it has spots.".

You are then given some examples labeled A to H. Given that you know that the general rule is true and you know that "Fish A is not green." is true, you have to say whether the underlined sentence "Fish A has spots." is true, YES, NO or MAYBE. So using the general rule "In

John's fishtank, if a fish is green it has spots." and knowing that "Fish A is not green." you have to say whether you think the sentence "Fish A has spots." must be true, cannot be true or might be true. If you think the sentence must be true you answer YES, if you think it cannot be true you answer NO and if you think it might be true or it might not be true, there is not enough information to say, you answer MAYBE."

The experimenter read through the examples from the first problem with the children asking them to mark their responses on the test booklet as the examples were read. Most subjects quickly understood what was required and preferred to work at their own pace. However they were told to wait for further instructions after completing the two problems from any session.

When all subjects had answered the two problems from session 1 they were handed out the appropriate pictures of the problem content for the empirical reasoning problems of session 2. The subjects were told that the picture demonstrated that the general rule was true. They were then asked whether they agreed that the general rule was true with respect to the relevant picture. Most subjects quickly agreed that it was although some argued that certain elements in the picture were counterexamples to the truth of the rule. When this happened the experimenter explained that the particular elements were in fact compatible with the truth of the rule and did not actually make the rule false. The subjects appeared to understand and accept this. The subjects were told that they had to answer the examples from the problems in session 1 again but this time they were to use the picture to help them. The subjects were told that they might find that their responses to the session 2 examples differed from their responses to the session 1 examples even although the problems were the same, since using the pictures would probably make the problems easier to answer. However the subjects were instructed not to look back at their session 1 responses. The experimenter then read through the examples of the first problem of session 2 along with the children who then carried on to complete the second problem of session 2.

When the subjects had answered both session 2 problems the pictures were taken away. Subjects were then instructed to answer the session 3 problems. They were told that, as in session 1, they had to answer the problems without a picture to help. However in responding to the session 3 examples they were asked to try to apply any insight that they had gained from answering the session 2 examples.

Subjects were told that the content of the problems in session 4 was different from that of the previous sessions and that once again the problems were to be answered without the help of a picture.

The experimenter remained present throughout the experiment in case of any difficulties but even the youngest subjects seemed to understand the instructions.

RESULTS AND DISCUSSION

SESSION 1

Since session 1 was prior to the training session performance on this session was predicted to be comparable to that found in previous class and conditional reasoning studies. Table 7.3 shows percent correct response for the different argument forms

TABLE 7.3

PERCENT CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS FOR GRADE AND SESSION

SESSION 1

	P5	P7	S2	S4	S6	TOT
pq	92.5	97.5	97.5	100	97.5	97
p-q	87.5	87.5	90	85	90	88
-q-p	75	72.5	82.5	95	82.5	81.5
-qp	82.5	92.5	85	95	92.5	89.5
-pq	27.5	17.5	27.5	25	50	29.5
-p-q	17.5	22.5	20	27.5	35	24.5
qp	10	15	7.5	17.5	45	19
q-p	20	25	10	30	37.5	24.5
TOT	51.6	53.8	52.5	59.38	66.3	56.7

SESSION 2

	P5	P7	S2	S4	S6	TOT
pq	97.5	92.5	87.5	92.5	92.5	92.5
p-q	85	90	95	92.5	92.5	91
-q-p	77.5	70	75	85	77.5	77
-qp	85	92.5	82.5	82.5	95	85.5
-pq	60	80	67.5	67.5	80	74.5
-p-q	58	80	85	67.5	80	60
qp	42.5	70	62.5	60	85	64
q-p	52.5	72.5	65	57.5	82.5	66
TOT	67.8	79.1	77.8	72.5	85.6	76.6

SESSION 3

	P5	P7	S2	S4	S6	TOT
pq	95	97.5	90	97.5	90	94
p-q	82.5	82.5	80	92.5	90	85.5
-q-p	72.5	52.5	42.5	85	82.5	67
-qp	77.5	82.5	72.5	87.5	77.5	79.5
-pq	42.5	72.5	82.5	60	87.5	69
-p-q	30	62.5	67.5	52.5	77.5	58
qp	30	32.5	40	35	67.5	41
q-p	42.5	55	70	60	70	59.5
TOT	59.1	67.2	68.1	71.3	80.3	69.2

TABLE 7.3 (CONT.)

SESSION 4

	P5	P7	S2	S4	S6	TOT
pq	95	97.5	87.5	92.5	85	91.5
p-q	70	85	62.5	80	77.5	75
-q-p	72.5	52.5	40	70	77.5	62.5
-qp	70	77.5	55	72.5	80	71
-pq	25	67.5	57.5	42.5	70	52.5
-p-q	32.5	52.5	50	30	65	46
qp	17.5	35	27.5	25	65	34
q-p	35	62.5	50	40	65	50.5
TOT	52.2	66.3	53.8	56.6	73.1	60.4

for subjects at different grades (also shown are percent correct responses on sessions 2, 3 and 4). A 5 (grade) X 2 (linguistic form) X 8 (argument form) analysis of variance was carried out on correct response and Table 7.4 shows the results of this analysis. Since neither linguistic form nor any of the interactions of linguistic form with other factors were significant a further analysis was carried out on the data excluding linguistic form in order to simplify tests of simple main effects. The results are shown in Table 7.4a. The main effect of grade was significant, $F(4,95)=2.95$, $p<0.05$. Although S4 and S6 subjects performed better than P5, P7 and S1 subjects Newman Keuls tests on comparisons between means for grade (Table 7.4b) showed that only the comparisons of the 6th. year subjects (66.3% correct) with P5 (51.6% correct), P7 (53.8% correct) and S2 (52.5% correct) were significant.

The highly significant main effect of argument form, $F(7,665)=161.39$, $p<0.001$, was largely due to the good performance on determinate argument forms, for which 89% of responses overall were correct, compared with the poor perfomance on indeterminate argument forms for which only 24% of responses were correct. Newman Keuls test on comparisons between argument form means (Table 7.4c) showed that, as expected, all comparisons between valid and invalid argument forms were highly significant. In addition comparisons between valid argument forms pq (Modus Ponens) and -q-p (Modus Tollens) and between invalid argument forms -pq and qp were significant.

The absence of a grade by argument form interaction contrasted with experiment 6 and the results of Kuhn (1977) and Taplin et al (1974) in showing that there was no differential improvement in performance for invalid compared with valid argument forms with increasing age.

TABLE 7.4

SESSION 1: ANALYSIS OF VARIANCE ON CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	392.844	1599			
BETWEEN	44.657	99			
GRADE	4.835	4	1.209	2.88	0.05
ERROR	39.822	95	0.419		
WITHIN	348.187	1500			
LING. FORM	0.455	1	0.455	3.55	0.1
LF X G	0.685	4	0.171	1.34	ns
ERROR(GL)	12.172	95	0.128		
ARGFORM	170.589	7	24.370	161.39	0.001
AF X G	5.545	28	0.198	1.31	ns
ERROR(AG)	100.553	665	0.151		
AF X LF	1.05	7	0.150	1.79	ns
AF X LF X G	3.235	28	0.115	1.37	ns
ERROR	55.953	665	0.084		

TABLE 7.4a

ANALYSIS OF VARIANCE ON CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS ACROSS GRADE ON SESSION 1

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	642.689	799			
BETWEEN	89.314	99			
GRADE	9.670	4	22.418	2.95	0.05
ERROR	79.644	95	0.838		
WITHIN	553.375	700			
ARG. FORM	341.179	7	48.740	161.39	0.001
A X G	11.09	28	0.396	1.31	ns
ERROR	201.106	665	0.302		

TABLE 7.4b

SESSION 1: NEWMAN KEULS COMPARISONS BETWEEEN GRADE MEANS

	P5 1.031	S2 1.050	P7 1.075	S4 1.188	S6 1.325	r	sq(0.95)(r,95) / sq(0.99)(r,95)
P5 1.031	-	0.019 ns	0.044 ns	0.157 ns	0.294 *	5	0.285 / 0.342
S2 1.050		-	0.025 ns	0.138 ns	0.275 *	4	0.268 / 0.328
P7 1.075			-	0.113 ns	0.250 *	3	0.243 / 0.305
S4 1.188				-	0.137 ns	2	0.202 / 0.269

TABLE 7.4c

SESSION 1: NEWMAN KEULS COMPARISONS BETWEEN ARGUNT FORM MEANS

	PQ 1.94	-QP 1.79	P-Q 1.76	-Q-P 1.63	-PQ 0.59	-P-Q/Q-P 0.49	QP 0.38	r	s q(0.95)(r,665) s q(0.99)(r,95)
PQ 1.94	-	0.15	0.18 ns	0.31 **	1.35 **	1.45 **	1.56 **	7	0.231 0.272
-QP 1.79		-	0.03	0.16 ns	1.20 **	1.30 **	1.41 **	6	0.223 0.265
P-Q 1.76			-	0.13 ns	1.17 **	1.27 **	1.25 **	5	0.214 0.256
-Q-P 1.63				-	1.04 **	1.14 **	1.25 **	4	0.201 0.245
-PQ 0.59					-	0.10 ns	0.21 *	3	0.183 0.229
-P-Q/Q-P 0.49						-	0.11 ns	2	0.153 0.202

In order to establish whether subjects responded according to a particular interpretation of the major premise the data for session 1 were also analysed in terms of correspondence of the responses by individual subjects to model response patterns. The model response patterns for different interpretations of the rule are shown in Table 7.5. When affirmative and negative versions of the arguments are used conjunctive and biconditional responses can be distinguished, as was mentioned in Chapter 6. However, in contrast to Taplin et al, very few conjunctive response patterns were found and consequently only conditional and biconditional response patterns were considered in the analysis. A criterion of 7 responses out of 8 correct according to a particular model pattern was used as a measure of response according to a particular pattern since this criterion determined a unique response pattern.

The classification of response patterns (Table 7.6) complemented the argument form analysis in showing that the low levels of correct response on the indeterminate argument forms were largely attributable to subjects adopting the biconditional response pattern: 60.5% of responses overall were biconditional compared with only 15.5% conditional and the remainder (24%) inconsistent. The biconditional interpretation was predominant at all grades from P5 through to 6th year secondary with a peak at second year secondary (75% biconditional response patterns).

The procedure of Marcus and Rips (1979) was used in order to establish whether the distribution of response patterns changed across grade. Conditional responses were scored as +1, biconditional as -1 and unclassifiable as 0. The resulting figures were subjected to an analysis of variance the results of which are shown in Table 7.7. The

TABLE 7.5

POSSIBLE MODEL RESPONSE PATTERNS FOR DIFFERENT INTERPRETATIONS OF THE CONDITIONAL

ARGUMENT FORM	CONJ.	BICOND.	COND.
PQ	T	T	T
P-Q	F	F	F
-PQ	F	F	M
-P-Q	F	T	M
QP	T	T	M
Q-P	F	F	M
-QP	F	F	F
-Q-P	F	T	T

TABLE 7.6

RESPONSE PATTERNS FOR SESSIONS 1 TO 4 ACROSS GRADE

u = unclassifiable
(or inconsistent)

SESSION 1

	CLASS			CONDITIONAL			TOTAL		
	B	C	U	B	C	U	B	C	U
P5	14	1	5	11	1	8	25	2	13
P7	12	5	3	14	1	5	26	6	8
S2	15	1	4	15	1	4	30	2	8
S4	10	5	5	13	3	4	23	8	9
S6	7	7	6	10	6	4	17	13	10
TOT	58	19	23	63	12	25	121	31	48

SESSION 2

P5	4	7	9	2	7	11	6	14	20
P7	2	10	8	1	11	8	3	21	16
S2	0	8	12	0	12	8	0	20	20
S4	1	7	12	1	6	13	2	13	25
S6	0	12	8	0	15	5	0	27	13
TOT	7	44	49	4	51	45	11	95	94

SESSION 3

P5	8	6	6	7	3	10	15	9	16
P7	4	5	11	3	6	11	7	11	22
S2	0	6	14	0	8	12	0	14	26
S4	5	9	6	5	7	8	10	16	14
S6	1	13	6	1	9	10	2	22	16
TOT	18	39	43	16	33	51	34	72	94

SESSION 4

P5	8	3	9	9	3	8	17	6	17
P7	3	8	9	4	7	9	7	15	18
S2	3	5	12	4	3	13	7	8	25
S4	9	4	7	6	5	9	15	9	16
S6	3	11	6	1	13	6	4	24	12
TOT	26	31	43	24	31	45	50	62	88

effect of grade was significant, $F(4,95)=2.57$, $p<0.05$ showing that the distribution of response patterns did change with grade but Newman Keuls comparisons showed that only the comparison between S2 and S6 was significant (Table 7.7a). It should be pointed out however that other methods of scoring might lead to more significant comparisons between grades. A chi-square test on distribution of response patterns between grade, for instance gives a significant P3/S6 comparison $\chi^2(2)=9.98$, $p<0.01$.

Comparison with other studies

In order to make a fair comparison between experiments 6 and 7 responses of P4 to S5 subjects from experiment 6 were compared with the responses of the P5 to S6 subjects in experiment 7. 57% of session 1 responses were correct in experiment 7 compared with 59% of responses in experiment 6. Overall more responses to valid argument were correct in session 1 (89%) than in experiment 6 (77.5%) and fewer responses to invalid argument forms (24.4% compared with 31.5%). There was a larger increase in correct response across grade in experiment 6 (42% to 74%), than experiment 7 (52% to 66%) which was not due to a wider age range since the age range being considered was the same although the subjects in experiment 6 were from P4 to S5 while those in experiment 7 were P5 to S6. The steeper rise in correct response with increasing age in experiment 6 seemed to be due to the very poor performance in experiment 6 by primary children on invalid argument forms (only 9% of P4 and 26.5% of P6 responses were correct), which improved dramatically with increasing age up to 71.5% correct at S5 compared with the relatively gradual increase in session 1 from 18.8% at P5 to only 41.9% correct at S6. It is possible that the advantage

TABLE 7.7

SESSION 1: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	170.98	99			
GRADE	16.70	4	4.175	2.57	0.05
ERROR	154.28	95	1.624		

TABLE 7.7a

SESSION 1: COMPARISONS BETWEEN GRADE MEANS FOR RESPONSE PATTERN ANALYSIS

GRADE ORDERED MEANS	S2 -1.40	P5 -1.15	P7 -1.00	S4 -0.75	S6 -0.20	r	s q(0.95)(r,95) s q(0.99)(r,95)
S2 -1.40	-	0.25	0.40	0.65 ns	1.20 *	5	1.13 1.32
P5 -1.15		-	0.15	0.40	0.95 ns	4	1.06 1.27
P7 -1.00			-	0.25	0.80	3	0.96 1.18
S4 -0.75				-	0.55	2	0.80 1.05

of the older subjects but not the younger subjects on invalid argument forms in experiment 6 compared with experiment 7 was due to some benefit derived by the older subjects from preceding the syllogism task by the evaluation task.

In terms of response patterns the percentage of conditional responses by 6th. year secondary subjects (33%) was comparable to that found for 5th. year secondary subjects in experiment 6 (34%) and is also comparable to that found for adults in Marcus and Rips study (32.3%) although 6th year subjects gave more biconditional responses (43%) than Marcus and Rip's subjects (14%) or 5th year secondary subjects in experiment 6 (13%). The larger number of biconditional response patterns reflected the relatively poor performance on the indeterminate argument forms in session 1 compared with experiment 6. It is probable that the complexity introduced by the inclusion of both affirmative and negative versions of the arguments in experiment 7 makes many subjects resort to the immature biconditional response pattern. It is also possible that the content of the problems in experiment 7 was more conducive than that in experiment 6 to the biconditional interpretation. Marcus and Rips found that problem content can significantly influence the response pattern adopted in a syllogistic reasoning task. They found that the incidence of biconditional response varied from 6.5% to 25% depending on the problem content. A further possible reason for the difference between experiments 6 and 7 in the incidence of biconditional response has already been suggested - the fact that the syllogism task in experiment 6 is preceded by the evaluation task possibly makes the subjects in experiment 6 more aware of the inadequacies of the biconditional response.

SESSION 2

RESULTS AND DISCUSSION

Percent correct response to the different argument forms in session 2 for class and conditional premises for subjects at different grades are shown in Table 7.3. The results of a 5 (grade) X 2 (linguistic form) X 8 (argument type) analysis of variance are shown in Table 7.8. As in session 1 there was no effect of linguistic form and neither were any interactions of linguistic form significant. In order to simplify the tests of simple effects a further analysis was performed excluding the effects of linguistic form (see Table 7.8a). Significant main effects were found for grade, $F(4, 95)=2.79$, $p<0.05$ and argument form, $F(7,665)=21.60$, $p<0.001$ and the grade by argument form interaction was significant $F(28, 665)=1.91$, $p<0.01$. Newman Keuls comparisons between grades revealed significant differences between means only for the comparison between P5 and S6 (Table 7.8b). The significant grade X argument form interaction indicated that the levels of correct response to different argument forms changed across grade. The simple main effect of argument form was significant at all grades (Table 7.8c) except 6th. year secondary indicating that only for subjects at this grade was there no difference in correct response on determinate and indeterminate argument forms. Simple main effects analysis of grade showed that performance on determinate argument forms did not change significantly across grade (Table 7.8d) while performance on all indeterminate argument forms except -pq did change across grade.

The classification of session 2 response patterns (Table 7.6) showed

TABLE 7.8

SESSION 2: ANALYSIS OF VARIANCE ON CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	287.109	1599			
BETWEEN	55.672	99			
GRADE	5.856	4	1.464	2.79	0.05
ERROR	49.816	95	0.524		
WITHIN	231.437	1500			
LING. FORM	0.390	1	0.390	2.96	ns
LF X G	0.107	4	0.027	0.21	ns
ERROR(GL)	12.565	95	0.132		
ARGFORM	22.604	7	3.229	21.67	0.001
AF X G	7.944	28	0.284	1.91	ns
ERROR(AG)	99.389	665	0.149		
AF X LF	1.125	7	0.161	1.27	ns
AF X LF X G	3.033	28	0.108	0.85	ns
ERROR	84.28	665	0.127		

TABLE 7.8a

SESSION 2: ANALYSIS OF CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	371.220	799			
BETWEEN	111.345	99			
GRADE	11.710	4	2.928	2.79	0.05
ERROR	99.635	95	1.049		
WITHIN	259.875	700			
ARG. FORM	45.210	7	6.459	21.60	0.001
AF X G	15.950	28	0.570	1.91	0.01
ERROR	8.715	665	0.299		

TABLE 7.8b

SESSION 2: NEWMAN KEULS COMPARISONS BETWEEEN GRADE MEANS

	P5 1.356	S2 1.450	P7 1.556	S4 1.581	S6 1.713	r	sq(0.95)(r,95) / sq(0.99)(r,95)
P5 1.356	-	0.094 ns	0.200 ns	0.225 ns	0.357 *	5	0.321 0.385
S4 1.450		-	0.106 ns	0.131 ns	0.263 ns	4	0.301 0.369
P7 1.556			-	0.025 ns	0.157 ns	3	0.274 0.343
S4 1.581				-	0.132 ns	2	0.228 0.302

TABLE 7.8c

SESSION 2: SUMMARY TABLE OF SIMPLE EFFECTS OF ARGUMENT FORM

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P5	25.144	7	3.592	12.01	0.001
P7	7.094	7	1.013	3.39	0.01
S2	7.444	7	1.063	3.56	0.01
S4	18.900	7	2.700	9.03	0.001
S6	2.575	7	0.368	1.23	ns
ERROR	198.715	665	0.299		

TABLE 7.8d

SUMMARY TABLE OF SIMPLE EFFECTS OF GRADE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
PQ	0.400	4	0.100	0.25	ns
P-Q	0.460	4	0.115	0.29	ns
-Q-P	0.940	4	0.235	0.60	ns
-QP	1.100	4	0.275	0.70	ns
-PQ	3.44	4	0.86	2.19	0.1
-P-Q	9.1	4	2.275	5.79	0.001
QP	7.66	4	1.915	4.87	0.001
Q-P	4.56	4	1.14	2.90	0.025
ERROR	298.35	760	0.393		

that in contrast with session 1 there were very few biconditional response patterns (only 5.5% overall) and this was true for subjects at all grades. The predominant response pattern was the conditional response pattern (47.5%) although there were almost as many unclassifiable response patterns (47%). As in session 1 an analysis of variance was carried out on the response pattern data in session 2 (Table 7.9). The significant main effect of grade, $F(4,95)=3.28$, $p<0.05$, showed that there was a significant difference in the distribution of response patterns across grade. Newman Keuls comparisons between grades (Table 7.9a) revealed that only the comparisons between P5 and S6 subjects and between S4 and S6 subjects were significant reflecting the higher number of conditional responses on S6 (67.5%) compared with S4 (32.5%) and P2 (35%).

Since the training session was specifically designed to help subjects respond correctly on the invalid argument forms performance on session 1 was compared with that on session 2 in order to establish whether there was a significant change in performance from session 1 to session 2 only for invalid argument forms. As expected overall performance on the empirical task of session 2 (77% correct) was superior to performance on session 1 (57% correct). A 5 (grade) X 2 (session) X 8 (argform) analysis of variance was carried out in order to compare performance on sessions 1 and 2 (Table 7.10) Significant main effects were found for grade, $F(4,95)=4.27$, $p<0.005$, session, $F(1,45)=73.75$, $p<0.001$ and argument form, $F(7,665)=124.24$, $p<0.001$. Significant interactions were found for grade X argument form, $F(28, 665)=4.40$, $p<0.001$, and session X argument form, $F(28, 665)=1.22$, $p<0.001$.

The session by argument form interaction showed that the improvement

TABLE 7.9

SESSION 2: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	93.44	99			
GRADE	11.34	4	2.835	3.28	0.05
ERROR	82.10	95	0.864		

TABLE 7.9a

SESSION 2: BETWEEN GRADE COMPARISONS FOR RESPONSE PATTERN DATA

	P5 0.40	S4 0.55	P7 0.90	S2 1.00	S6 1.35	r	s q(0.95)(r,95) s q(0.95)(r,95)
P5 0.40	-	0.15	0.50	0.60 ns	0.95 *	5	0.822 0.946
S4 0.55		-	0.35	0.45 ns	0.80 *	4	0.772 0.946
P7 0.90			-	0.10 ns	0.45 ns	3	0.703 0.882
S2 1.00				-	0.35	2	0.584 0.776

TABLE 7.10

SESSION 1/2: ANALYSIS OF CORRECT RESPONSE FOR ARGUMENT FORM, GRADE AND SESSION

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	1077.110	1599			
BETWEEN	115.485	99			
GRADE	17.604	4	4.401	4.27	0.05
ERROR	97.881	95	1.030		
WITHIN	961.625	1500			
SESSION	63.203	1	63.203	73.75	0.001
S X G	3.778	4	0.945	1.10	ns
ERROR(SG)	81.394	95	0.857		
ARGFORM	307.870	7	43.981	24.24	0.001
A X G	43.586	28	1.557	4.40	0.001
ERROR(AG)	235.169	665	0.354		
S X A	78.517	7	11.217	45.23	0.001
S X A X G	8.452	28	0.302	1.22	ns
ERROR(AG)	164.656	665	0.248		

TABLE 7.10a

SESSION 2: SUMMARY TABLE OF SIMPLE EFFECTS OF SESSION

SOURCE	SUM OF SQUARE	DEGREES OF FREEDOM	MEAN SQUARE	F	P
PQ	0.405	1	0.405	1.25	ns
P-Q	0.180	1	0.180	0.56	ns
-Q-P	0.405	1	0.405	1.25	ns
-QP	0.080	1	0.080	0.25	ns
-PQ	40.500	1	40.500	125.00	0.001
-P-Q	45.045	1	45.045	139.03	0.001
QP	42.155	1	42.155	130.11	0.001
Q-P	34.445	1	34.445	106.31	0.001
ERROR	246.050	760	0.324		

in performance from session 1 to session 2 was largely due to improvement in performance on invalid argument forms (from 24.4% correct on session 1 to 66.1% correct on session 2). Although correct response on valid argument forms actually decreased from session 1 (89% correct) to session 2 (86.5% correct) tests of simple effects of session at argument form(i) (Table 7.10a) showed that only the increases in correct response on invalid argument forms were significant. The grade X argument form interaction reflected the fact that the differentiation in response to valid and invalid argument forms was greater at some grades than others.

In order to compare the distribution of response patterns on sessions 1 and 2 a 5 (grade) X 2 (session) analysis of variance was carried out on the response pattern data with grade as a between subjects factor and session as a within subjects factor. As Table 7.11 shows significant main effects were found for both grade, $F(4,95)=3.14$, $p<0.05$, and session $F(1,95)=181.29$, $p<0.001$, but the grade by session interaction was not significant. The significant effect of session reflected a change in response from a predominantly biconditional response on session 1 (61% biconditional), with only 15.5% correct conditional response, to predominantly conditional response on session 2 (47.5% conditional responses) with only 5.5% biconditional.

The absence of a significant grade X session interaction in both the argument form and response pattern analyses indicated that the improvement in performance from session 1 to session 2 was similar across all grades. The significant effect of argument form in the session 2 analysis reflected the fact that even in the concretised task invalid argument forms are more difficult to evaluate than valid argument forms although there are exceptions to this. If the absence

TABLE 7.11

SESSION 1/2: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	415.82	199			
BETWEEN	177.82	99			
GRADE	20.77	4	5.19	3.14	0.05
ERROR	157.05	95	1.65		
WITHIN	238.00	100			
SESSION	151.38	1	151.38	181.29	0.001
S X G	7.27	4	1.82	2.18	0.1
ERROR	79.35	95	0.84		

of a significant effect of argument form is taken as an index of mature performance only 6th. year subjects on session 2 performed maturely on the task. Insofar as response on the invalid argument forms on session 2 was superior to that on session 1 for subjects at all grades the training task was potentially effective.

In view of the significant effects of linguistic form in the verification tasks which are also empirical tasks the absence of an effect of linguistic form in the "empirical reasoning" task of session 2 is perhaps surprising. However consideration of the "empirical reasoning" task shows that it is not actually necessary in solving this task to consider the major premise at all and the task could in fact be solved in the absence of the major premise. The major premise is simply a true description of the empirical situation although presumably if the description were false subjects might notice the contradiction. Successful solution of the empirical reasoning task depends upon the subject being able to establish whether the minor premise identifies a unique set of elements or not. In the former case the inference is valid and in the latter case invalid. Successful solution of the syllogistic reasoning task would require that the insight gained in the concrete task transferred to the abstract linguistic condition. This insight would be that correct response depends upon whether the minor premise determines a unique type of element and the realisation that this information can be recovered from the relationship expressed in the major premise. The predictions of a competence model would be that the insight would transfer only for those subjects who had the competence to understand the inferential properties of the implication relationship.

SESSION 3

RESULTS AND DISCUSSION

Performance on session 3 was to be compared with performance on session 1 in order to establish whether the intervening "training session" was effective in promoting correct response on invalid argument forms on the linguistic reasoning task. Percent correct response to each argument form for subjects at different grades on Session 3 are shown in Table 7.3. As in the previous sessions preliminary analysis of correct response on session 3 revealed that neither the main effect of linguistic form nor any of the interactions of linguistic form were significant and consequently this factor was excluded from further analyses. A 5 (grade) X 2 (session) X 8 (argument form) analysis of variance was carried out on correct response on sessions 1 and 3 (Table 7.12). Significant main effects were found for grade, $F(4,95)=3.86$, $p<0.01$, session, $F(1,95)=50.81$, $p<0.001$, and argument form, $F(7,665)=122.09$, $p<0.001$. Significant interactions were found for grade X argument form, $F(28,665)=2.29$, $p<0.05$, argument form X session, $F(7,665)=33.69$, $p<0.001$, and grade X session X argument form, $F(28,665)=1.94$, $p<0.01$.

The main effect of session showed that performance on session 3 was significantly better than performance on session 1 with 69.2% of session 3 responses correct overall compared with only 56.7% of session 1 responses. Although the increase in correct response from session 1 to session 3 was greater at some grades than others (the smallest increase was 7.5% for primary 5 subjects and the largest 14% for primary 7 subjects) the F ratio associated with the session X grade interaction indicated that the difference in the magnitude of

TABLE 7.12

SESSION 1/3: ANALYSIS OF VARIANCE ON CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS ON SESSIONS 1 AND 3

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
TOTAL	1174.877	1599			
BETWEEN	168.127	99			
GRADE	23.521	4	5.880	3.86	0.01
ERROR	144.606	95	1.522		
WITHIN	1006.750	1500			
SESSION	25.000	1	25.000	50.81	0.1
S X G	1.231	4	0.308	0.63	ns
ERROR(SG)	46.769	95	0.492		
ARGFORM	352.097	7	50.300	122.09	0.001
AF X G	26.409	28	0.943	2.29	0.05
ERROR(AG)	274.244	665	0.412		
AF X S	69.340	7	9.906	33.69	0.001
AF X S X G	15.979	28	0.571	1.94	0.01
ERROR(ASG)	195.661	665	0.294		

TABLE 7.12a

SESSION 3: SUMMARY TABLE OF SIMPLE EFFECTS OF SESSION

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
PQ	0.180	1	0.180	0.56	ns
P-Q	0.125	1	0.125	0.39	ns
-Q-P	4.205	1	4.205	13.18	0.001
-QP	2.000	1	2.000	6.27	0.025
-PQ	31.205	1	31.205	97.82	0.001
-P-Q	22.445	1	22.445	70.36	0.001
QP	9.680	1	9.680	30.34	0.001
Q-P	24.500	1	24.500	76.80	0.001
ERROR	242.430	760	0.319		

TABLE 7.12b

SESSION 1/3: SUMMARY TABLE OF SIMPLE ARGUMENT FORM X SESSION INTERACTION AT GRADE(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P5	3.600	7	0.514	1.75	ns
P7	19.647	1	2.807	9.55	0.001
S2	14.737	1	2.105	7.16	0.001
S4	8.588	1	1.227	4.17	0.001
S6	13.897	1	1.985	6.75	0.001
ERROR	195.661	665	0.294		

the increase between sessions for different grades was not significant.

The argument form X session interaction indicated that the change in performance from session 1 to session 3 was not the same for all argument forms. In fact examination of Table 7.3 shows that there was not only an increase in correct response on invalid argument forms (i.e. affirmative and negative versions of DA and AC) from session 1 (24.5% correct) to session 3 (56.9% correct) but there was also a decrease in correct response on determinate argument forms from session 1 (89% correct) to session 3 (81.5% correct). Tests on simple effects of session for the session X argument form interaction showed that (Table 12a) the increases in correct response on all invalid argument forms were significant as were the decreases in correct response on both affirmative and negative versions of Modus Tollens. The session X argument form X grade interaction indicated that the changes in response on the different argument forms from session 1 to session 3 were not the same for all grades. An analysis of simple interaction effects showed that the session X argument form interaction was significant at all grades except primary 5 (Table 7.12b).

The distribution of response patterns for session 3 (Table 7.6) showed that the predominant response pattern was unclassifiable: overall 47% of response patterns were unclassifiable with 18% of response patterns biconditional and 35% conditional. An analysis of change in response pattern across grade on session 3 (Table 7.13) revealed a significant main effect of grade, $F(4,95)=3.26$, $p<0.025$, reflecting a general increase in correct conditional response with grade from 22.5% at primary 5 to 55% at 6th. year secondary and a variation in the prevalence of biconditional response across grade: 37.5% of P5

TABLE 7.13

SESSION 3: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	163.56	99			
GRADE	19.76	4	4.94	3.26	0.025
ERROR	143.80	95	1.51		

TABLE 7.13a

SESSION 3: BETWEEN GRADE COMPARISONS FOR RESPONSE PATTERN DATA

	P5 -0.30	P7 0.20	S4 0.30	S2 0.70	S6 1.00	r	s q(0.95)(r,95) s q(0.99)(r,95)
P5 -0.30	-	0.50	0.60	1.00 ns	1.30 *	5	1.086 1.309
P7 0.20		-	0.10	0.50 ns	0.80 ns	4	1.023 1.249
S4 0.30			-	0.40	0.70	3	0.930 1.166
S2 0.70				-	0.30	2	0.773 1.023

response patterns were biconditional with 0% at S2 but 25% at S4. Newman Keuls tests on comparisons between grade means (Table 7.13a) showed that only the comparison between Primary 5 and 6th. year secondary was significant. The significance of the P5/S6 comparison reflected the large decrease in biconditional and increase in conditional response pattens from P5 to S6 and the same was true for the P5/S2 comparison.

In order to compare distributions of response patterns between sessions 1 and 3 a 5 (grade) X 2 (session) analysis of variance was carried out on response patterns (Table 7.14). Significant main effects were found for grade, $F(4,95)=3.21$, $p<0.025$, and session $F(1,95)=93.41$, $p<0.001$, and the grade X session interaction was significant, $F(4,95)=6.49$, $p<0.001$.

Simple effects analysis showed that session was significant at all grades (Table 7.14a). The significance of the difference in the distribution of response patterns between S1 and S3 for P5 subjects contrasts with the non-significance of the session x argument form interaction for P5 subjects.

The significant main effect of session showed that there was a net increase in correct response from session 1 to session 3. This indicated that the training session was moderately successful in facilitating correct response on the invalid argument forms although the overgeneralisation of the indeterminate response to Modus Tollens detracted from the success of the training. The non-significance of the grade X session interaction indicated that the effectivenes of training held for subjects at all grades although the non-significance of the simple interaction of session X argument form at primary 5 indicated that for subjects at this grade the magnitude of the

TABLE 7.14

SESSION 1/3: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	416.48	199			
BETWEEN	22.48	99			
GRADE	27.23	4	6.807	3.21	0.05
ERROR	201.25	95	2.118		
WITHIN	188	100			
SESSION	81.92	1	81.9	93.41	0.001
S X G	22.75	4	5.688	6.49	0.001
ERROR	83.33	95	0.877		

TABLE 7.14a

SESSION 1/3: SUMMARY TABLE OF SIMPLE EFFECTS OF SESSION OM RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P5	7.225	1	7.225	8.24	0.001
P7	14.400	1	14.400	16.42	0.001
S2	44.100	1	44.100	50.29	0.001
S4	11.025	1	11.025	12.57	0.001
S6	14.400	1	14.400	16.42	0.001
ERROR	83.33	95	0.877		

increase in correct response on invalid argument forms was not as large as at other grades. The response pattern data supported the demonstration of the effectivenes of training found for the argument form data in showing a significant reduction in the incidence of biconditional response and an increase in conditional response but also in inconsistent response.

SESSION 4

RESULTS AND DISCUSSION

In order to assess whether the effectiveness of training generalised to problems from a different content area performance on session 1 was compared with that on session 4. Percent correct response for each argument form for subjects at each grade on session 4 are shown in Table 7.3. A 5 (grade) X 2 (session) X 8 (argument form) analysis of variance was carried out on correct response on sessions 1 and 4 (Table 7.15). Significant main effects were found for grade $F(4,95)=3.48$, $p<0.05$, and argument form, $F(7,665)=138.77$, $p<0.001$. The argument form X grade interaction, $F(28, 665)=2.22$, $p<0.005$, was significant as was the argument form X session interaction, $F(7, 665)=34.38$, $p<0.001$, and the argument form X session X grade interaction, $F(28, 665)=2.42$, $p<0.001$.

Overall only 60.4% of session 4 responses were correct compared with 56.7% on session 1 and the main effect of session did not quite reach significance, $F(1,95)=3.92$, $p<0.1$. There was very little difference in correct response between sessions 1 and 4 for subjects at P5, S2 and S4 (S4 subjects actually made fewer correct responses overall on session 4 (56.6% correct) than session 1 (59.4% correct) but P7 and S6 subjects

TABLE 7.15

SESSION 1/4: ANALYSIS OF VARIANCE ON CORRECT RESPONSE TO DIFFERENT ARGUMENT FORMS ON SESSIONS 1 AND 4 ACROSS GRADE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	1143.410	1599			
BETWEEN	200.732	99			
GRADE	25.660	4	6.415	3.48	0.05
ERROR	175.072	95	1.843		
WITHIN	942.687	1500			
SESSION	2.175	1	2.175	3.92	0.1
S X G	4.653	4	1.163	2.10	0.1
ERROR(SG)	52.734	95	0.555		
ARGFORM	382.714	7	54.674	138.77	0.001
AF X G	24.520	28	0.876	2.22	0.005
ERROR(AG)	261.953	665	0.394		
AF X S	52.940	7	7.563	34.38	0.001
AF X S X G	14.907	28	0.532	2.42	0.001
ERROR(ASG)	146.091	665	0.220		

TABLE 7.15a

SESSION 4: SUMMARY TABLE OF SIMPLE EFFECTS OF SESSION

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
PQ	0.605	1	0.605	2.31	ns
P-Q	3.380	1	3.380	12.90	0.001
-Q-P	7.220	1	7.220	27.56	0.001
-QP	6.845	1	6.845	26.13	0.001
-PQ	10.580	1	10.580	40.38	0.001
-P-Q	9.245	1	9.245	35.29	0.001
QP	4.500	1	4.500	17.18	0.001
Q-P	13.520	1	13.520	51.60	0.001
ERROR	198.825	760	0.262		

TABLE 7.15b

SESSION 1/4: SUMMARY TABLE OF SIMPLE ARGUMENT FORM X SESSION INTERACTION AT GRADE(i)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F	P
P5	42.211	7	6.030	27.41	0.001
P7	18.350	7	6.030	11.91	0.001
S2	29.400	7	4.200	19.09	0.001
S4	6.472	7	0.926	4.21	0.001
S6	10.288	7	1.470	6.68	0.001
ERROR	146.091	665	0.220		

had more correct responses on session 4 than session 1. The session x grade interaction was not significant however.

The argument form X grade interaction reflected different levels of correct response to different argument forms at different grades. The argument form X session interaction as in the session 3 analysis, reflected not just a general increase in correct response on valid argument forms from session 1 (24.4% correct) to session 4 (45.7% correct) but also a general decrease in correct response on invalid argument forms from session 1 (89% correct) to session 4 (75% correct). Tests of simple effects of session on the argument form X session interaction revealed significant effects of session for all argument forms except the affirmative version of Modus Ponens (Table 7.15a). Performance on all invalid argument forms increased from session 1 to session 4 while that on the valid argument forms decreased from session 1 to session 4. The argument form X session X grade interaction indicated that the changes in response on different argument forms from session 1 to session 4 were not the same for subjects at all grades. An analysis of simple interaction effects showed that the session x argument form interaction was significant at all grades (Table 7.15b).

Table 7.6 shows the distribution of response patterns across grade for session 4. As in session 3 the predominant response was unclassifiable (44%). Of the consistent response patterns there were more conditional response patterns (31%) than biconditional (25%). An analysis of variance was carried out on session 4 response patterns (Table 7.16) and a significant effect of grade was found, $F(4,95)=4.49$, $p<0.005$. Newman Keuls tests on comparisons between grade means (Table 7.16a) showed that the comparisons between P5 and S6 and S4 and S6 were

TABLE 7.16

SESSION 4: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	186.56	99			
GRADE	29.66	4	7.415	4.49	0.005
ERROR	156.900	95	1.652		

TABLE 7.16a

SESSION 4: BETWEEN GRADE COMPARISONS FOR RESPONSE PATTERN DATA

	P5 -0.55	S4 -0.30	S2 0.05	P7 0.40	S6 1.00	r	s q(0.95)(r,95) s q(0.95)(r,95)
P5 -0.55	-	0.25	0.60	0.95 ns	1.55 **	5	1.134 1.366
S4 -0.30		-	0.35	0.70 ns	1.30 **	4	1.068 1.303
S2 0.05			-	0.45 ns	0.95 ns	3	0.970 1.217
P7 0.40				-	0.60	2	0.806 1.070

significant.

A 5 (grade) X 2 (session) analysis of variance was carried out on the response pattern data from sessions 1 and 4. The results of the analysis (Table 7.17) revealed significant main effects of grade, $F(4.95)=3.92$, $p<0.01$, and session $F(1,95)=60.00$, $p<0.001$ and a significant grade by session interaction, $F(4,95)=2.49$, $p<0.05$. An analysis of simple main effects of session revealed that the effect of session was significant at all grades except 4th. year secondary (Table 7.17a).

The non-significance of the main effect of session and the grade by session interaction indicated that the effectiveness of the training session did not generalise to performance on problems from a different content area. The significant overall improvement in performance on session 3 for all grades except primary 5 but the non-significant increase in correct response on session 4 suggested that the increase in correct response on session 3 was not attributable to a strictly formal analysis. However although the training session did not lead to a significant net increase in correct response the training session did influence response given on session 4 by increasing the number of MAYBE responses used. However the significant increase found in correct response on invalid argument forms was accompanied by a corresponding decrease in correct response on all valid argument forms except the affirmative version of Modus Ponens indicating that although the training task was effective in promoting the use of the MAYBE response it was not effective in promoting an understanding of the appropriateness of the MAYBE response. The significant effect of session at all grades except fourth year secondary reflected a decrease in biconditional response which was accompanied largely by an

TABLE 7.17

SESSION 1/4: ANALYSIS OF RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
TOTAL	409.58	199			
BETWEEN	266.58	99			
GRADE	37.73	4	9.432	3.92	0.01
ERROR	228.85	95	2.409		
WITHIN	143.00	100			
SESSION	52.02	1	52.020	60.00	0.001
S X G	8.63	4	2.158	2.49	0.05
ERROR	82.35	95	0.867		

TABLE 7.17a

SESSION 1/4: SUMMARY TABLE OF SIMPLE EFFECTS OF SESSION OM RESPONSE PATTERN DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	P
P5	3.600	1	3.600	4.15	0.05
P7	19.600	1	19.600	22.61	0.001
S2	21.025	1	21.025	24.25	0.001
S4	2.025	1	2.025	2.34	ns
S6	14.400	1	14.400	16.61	0.001
ERROR	82.35	95	0.867		

increase in inconsistent response but also by an increase in conditional response. This indicated that the intervening training had made the subjects aware of the inadequacy of the matching (biconditional) response pattern but, although there was an increase in conditional response patterns at most grades, the overgeneralisation of the MAYBE response to valid argument forms indicated that the training had not led to a better understanding of the inferential properties of the implication relationship and the asymmetry of the inclusion/implication relationship still caused difficulty. The unexpectedly poor performance by the fourth year secondary subjects was particularly difficult to explain but it seems likely that the results stemmed from the fact that the sample of fourth year secondary subjects contained a disproportionate number of non-certificate pupils.

The results can either be interpreted as evidence that adolescents and young adults do not have the formal competence to benefit from training or simply as evidence that the training task selected was ineffective as a training task. It is possible that the subjects were simply too young to benefit from the training. In O'Brien and Overton's studies (1980, 1982) only the subjects who were presumably well into the formal operational stage (mean ages 21 years 5 months and 18 years 2 months) respectively benefited from the contradiction training.

Studies using the selection task have shown a similar lack of transfer of ability to solve a concrete version of the selection task to the solution of an abstract task (Johnson-Laird, Legrenzi and Legrenzi, 1972; Van Duyne, 1974, 1976) although the selection task is intrinsically more difficult than the syllogistic reasoning task.

The non-significant effect of linguistic form in the session 2 analysis suggests one reason why the training was not effective in promoting an understanding of the distinction between valid and invalid argument forms. It was mentioned in the discussion of session 2 that the concrete task of session 2 could be answered in the absence of the major premise or at least without any consideration of the relationship between the major premise and the picture. Although at the beginning of the concrete task subjects agreed that the major premise was true of the picture the subjects probably did not appreciate the significance of the relationship between the picture and the inclusion/implication statement.

Since previous training tasks which have been successful in improving performance on conditional reasoning have demonstrated to the subject the contradiction between his initial inference and the correct inference the "concrete" training used in experiment 7 could probably be made more effective by emphasising the contradiction between the incorrect inferences made in the standard task and the correct inferences made in the concrete task. This could be achieved by having the subject respond to a particular argument form in the standard linguistic task and then having the subject respond to that argument form again immediately in the context of the concrete picture. If the subject's initial response was incorrect the contradiction between the incorrect response and the (presumably correct!) concrete response would possibly alert the subject to the structure of the inclusion/implication rule and consequently to correct response based upon an understanding of the formal properties of implication.

SUMMARY

In Chapter 7 an attempt was made to train subjects from 9 years to 17.5 years to respond correctly on the invalid argument forms D.A. and A.C. It was felt that a demonstration of the invalidity of DA and AC in a "concrete" version of a syllogistic reasoning task would alert the formally competent subject to the fact that some argument forms are indeterminate. In the training task subjects were required to respond to the syllogistic argument forms with the aid of a pictorial representation of the implication/class inclusion relation of the major premise. Performance on the eight forms of the standard categorical and conditional syllogistic reasoning tasks after training was compared with performance prior to training. The results showed significant increases in correct response overall at all grades on problems with similar content to the initial problem but no significant increase in overall correct response at any grade with problems of a different content. Performance on invalid forms did show a significant improvement but the indeterminate response was overgeneralised to the valid argument forms and a significant decrease in correct response was found on valid arguments. In conjunction with the response pattern analysis these results indicated that while the intervening training had made the subjects aware of the inadequacy of the matching (biconditional) response pattern which they had previously used, it had not promoted a better understanding of the inferential properties of the implication relationship. The lack of transfer from concrete to abstract problems is, however, consistent with results of similar studies using the selection task.

CHAPTER 8

SUMMARY AND DISCUSSION

This thesis has examined the development of reasoning abilities with class and conditional logic statements. According to Piaget's theory the acquisition of operational classification enables the concrete operational subject to understand quantified statements like "All A are B". In experiment 1 the verification paradigm was used to look at how subjects interpret universally quantified statements in an extensional context. A specific prediction of Bucci's (1977) structure neutral hypothesis that young subjects should perform better on a verification task with false than with true universally quantified propositions was also tested. The results supported Piaget's theory in showing good performance by 6 year olds on this task but the similar error rates found on true and false statements did not support Bucci's hypothesis.

Propositionally based models of verification which propose that verification is a function of propositional structure would predict that, since the general conditional "If x is an A then x is a B" also expresses universal quantification and is logically equivalent to the class statement "All A are B", concrete operational subjects should also perform well on a conditional verification task. Kuhn (1977) however claimed that conditional verification is a formal operational ability depending upon the subject's ability to interpret empirical evidence in terms of a propositional relationship.

In experiments 2 and 3 the class verification experiment was extended to include all possible class inclusion relations which can be made by combining positive and negative values of two binary attributes p and q and performance in verifying class statements was compared with performance in verifying conditional statements. Even the youngest subjects (primary 2) performed well on class verification although subjects still made some errors with certain types of class statement.

Performance on conditional verification was significantly worse than performance on class verification for subjects up to first year secondary. Primary school children had difficulty in verifying specific types of conditional statement but significant increases in correct response occurred at first year secondary. The pattern of errors on different statement types found for the young subjects indicated that they apparently understood the conditional as a conjunction. These results replicated Kuhn's results and were compatible with Kuhn's claim that conditional verification requires formal operational thinking.

Propositionally based models of verification could not account for the difference in response patterns on the different linguistic forms for the primary subjects. This differentiation in response on class and conditional verification was apparently compatible with Piaget's theory in showing that concrete operational subjects can understand class inclusion statements but not until formal operations are propositional logic statements understood. It seems that the young subjects can understand the class statements in the verification task because they map in an intuitively obvious way onto the underlying operations of class logic but the conditional statements do not and, in order to interpret conditional statements, the combinatorial structure of propositional logic is required.

Although the conditional verification task showed that the conditional interpretation of empirical information is apparently a formal operational ability it is possible that the difficulty that the young subjects have with the general conditional is caused by their inability to understand the general conditional as corresponding to the universal quantifier. It was predicted that if this was the source of the young subjects' difficulty in conditional verification, performance on the conditional verification task should be improved by presentation of the general conditional along with other universal quantifiers in a within subjects design. The results of experiment 4 showed that performance on conditional verification could be significantly improved by highlighting the interpretation of the general conditional as a universal quantifier in this way. However performance on conditional verification was still significantly poorer than performance on class verification indicating that there was still difficulty in mapping the general conditional onto the appropriate concrete operational classification operations.

Although these results were regarded as compatible with a certain interpretation of Piaget's theory difficulties were found in making clear predictions from Piaget's theory about performance on specific deductive reasoning tasks. It is not clear for instance from Piaget's use of the logic of classes and relations and propositional logic as models of concrete and formal operational intelligence respectively exactly what the distinction is between the logical abilities of subjects at these levels. Brainerd, for instance, argued that it does not make sense to use the more sophisticated logical system of classes and relations as a model of a less mature cognitive stage while Ennis argued that Piaget's "propositional" logic is really propositional functional logic which in modern systems of logic is not distinguished from class logic. From Ennis's interpretation of Piaget's logical systems no difference in performance on class and conditional logic problems would be predicted.

If the distinction in performance found on class and conditional verification problems is attributable to the use of different underlying cognitive operations in their solution this distinction in performance would also be expected on other logical reasoning tasks and it has been suggested for instance that correct response on class syllogistic reasoning problems should precede that on analogous conditional syllogisms (Roberge and Paulus, 1972). Subjects at the stage of concrete operations who can understand that "All A are B" is true even although there are not-as which are b and not-as which are not-b, as in the verification task, would be expected to understand that -a-b elements and -ab elements are compatible with the truth of a universally quantified statement. Although the ability of primary school subjects to verify class statements was attributed to their comprehension of the logic of inclusion it is likely that it was mediated simply by an understanding of the empirical inclusion relation. The ability to evaluate the truth of a contingent empirical relationship does not guarantee that the subject will understand the logical consequences of the inclusion relationship in the absence of appropriate concrete support. Indeed Piaget's use of a class inclusion example to exemplify implication, which caused problems for Ennis in trying to distinguish between the logical abilities of concrete and formal operational subjects, indicates that knowledge of the formal properties of inclusion requires formal operational thinking and inference tasks using inclusion statements will require knowledge of the logic of inclusion. Since "implication corresponds with inclusion" knowledge of the logic of inclusion presumably involves knowledge that inclusion corresponds with the combination (p.q)(-p.q)(-p.-q). It seems likely that it is not the kind of logic, class or propositional, but the depth of logical analysis required in the solution of a task which determines whether a subject can solve an inference task.

In experiments 5 and 6 performance on two inference tasks - an evaluation task and a syllogistic reasoning task - with class and conditional premises was examined. A modified version of the evaluation task including both class and conditional statements (rules), which was designed so that young children would be able to understand the task requirements, showed that there was no significant effect of linguistic form and primary school subjects performed poorly on the evaluation task with both class and conditional rules. The results of the evaluation task showed that when the inclusion/implication statement has no semantic or empirical support young subjects (under 12-13 years) were not able to use the logical structure of the rule to work out that different possible situations are compatible with the rule: those who did not accept only pq elements as compatible with the rule were just as likely to accept p-q elements as compatible with the rule as -pq and -p-q elements. Significant changes in the patterns of exemplars accepted were found between primary and secondary school with secondary subjects realising that -p-q elements were compatible with the rule although these subjects still had difficulty with the asymmetry of the rule, particularly with the conditional. The changes in the pattern of exemplars accepted as compatible with the rule at adolescence were regarded as consistent with Piaget's proposal that a qualitative change in the child's thinking occurs at adolescence with a "reversal in the direction between reality and possibility".

In experiment 6 subjects from 6 to 17 years were required to say whether conclusions to the four standard class or conditional syllogistic argument forms followed from the premises. As with the evaluation task there was no difference in correct response on class and conditional problems, indicating that understanding the logical consequences of both class statements and conditional statements requires formal operational thinking.

As in other studies the poor performance of younger subjects on the invalid argument forms DA and AC in experiment 6 contrasted with the good performance on the valid argument forms MP and MT. It was argued that the results of the syllogistic reasoning task, when considered in conjunction with the results of the evaluation task, indicated that the biconditional responses of the younger subjects were not logical biconditional responses since a logical biconditional response would be generated from an understanding that -p.-q elements are compatible with the major premise, but instead reflected a less mature "transductive" or "matching" response pattern. The increase in correct response on the evaluation task at adolescence was accompanied by an increase in correct response on the invalid arguments on the syllogistic reasoning task due to the realisation that the matching strategy was inadequate. Rappoport (Ennis, 1976), Staudenmayer and Bourne (1977) and others have suggested that correct response on the invalid argument forms is a critical test of a formal approach to reasoning problems and the increase in correct response on the invalid arguments indicated that the subjects were beginning to adopt a strategy which took cognisance of the formal properties of inclusion/implication. The decrease in correct response on MT at the same age also suggests a change in strategy albeit a change to a strategy which is not as successful on that argument form.

correct, it had not promoted a better understanding of the inferential properties of the inclusion/implication relationship since it had not enabled the subjects to establish the appropriateness of the MAYBE response. The results can be interpreted either as evidence that adolescents and adults do not have the formal competence to benefit from training or simply as showing the ineffectiveness of the training task. It seems likely that the training task did not bring out clearly enough the contradiction between the subjects' initial incorrect inferences and the correct inference.

The errors made by subjects even up to the age of 17-18 years on the syllogistic reasoning tasks and in the evaluation and conditional verification task indicated that adolescents and young adults were simply not as good at interpreting implication as a propositional relationship within the context of an integrated system of propositional relationships, differentiating it from other propositional relationships, testing the conditions under which it is true and establishing the logical consequences of implication as Piaget's theory would suggest.

It is useful to consider Falmagne's perspective on the development of reasoning ability with linguistic premises which was described in Chapter 1. Falmagne argued that there are two different methods of solving reasoning problems - those based on concrete representations of problems and those based on representations of the formal properties of the problem. However Falmagne proposed a less dramatic transition in reasoning ability from concrete-based to formal approaches to problems than that advocated by Piaget, arguing that one aspect of logical development is the change in predominance of different representational modes with age and particularly an increase in the availability of the formal mode of reasoning for solving problems with increasing age. Falmagne emphasises however that the availability of the formal representational mode does not guarantee its use since it may for example be more expedient to rely on concrete based modes of reasoning which have previously proved to be effective.

Falmagne regards the acquisition of logical competence as semantically based. Initially inferences are bound to a specific content but gradually, with repeated exposure to examples of these arguments in his "linguistic environment" and feedback about their validity, the subject comes to abstract the logical forms of arguments. Falmagne claims that the obvious implication of this is that problems with a familiar content will be dealt with at an earlier age than those with an unfamiliar content and indeed this is a consistent result in the literature. However whether such inferences with familiar content should be called deductive or whether they are really inductive is rather unclear. One prediction of Falmagne's account of the development of reasoning ability would be that problems which can be solved only by formal means will be particularly difficult to solve and the ability to solve such problems will appear relatively late. The content of the problems considered in this thesis was specifically chosen as expressing plausible contingent relationships. Indeed the results of the experiments reported in this thesis, which used problems expressing plausible, contingent relationships, and many other studies show that problems which can only be solved by formal methods alone still cause difficulty for older subjects.

Even Piaget (1972) did acknowledge that the subject is more likely to demonstrate his logical competence in a subject area with which he is familiar than in one with which he is not familiar. Although it was claimed earlier on that it seemed rather paradoxical to attempt to demonstrate formal competence by showing that a subject can reason formally with familiar content, Johnson-Laird (1983) points out, the early logicians proceeded in exactly that way in formulating the principles of valid inference i.e. by abstracting the principles of valid logical argument from consideration of semantically different instances of argument forms. This view of the development of formal competence suggests a specific view of the role of logic in explanations of reasoning ability i.e. that the principles of logic are normative with respect to reasoning, that they are derived as Cohen (1981) suggests from a sytematisation of intuitions about logical validity.

Although Johnson-Laird's mental model theory could account for the increasingly correct response patterns on the evaluation and syllogistic reasoning tasks in terms of the construction of increasingly accurate models of the premises it would not have predicted that the qualitative changes in reasoning would occur at adolesence and a Piagetian component would need to be invoked to account for this.

The general conclusion to the thesis is that the significant changes in patterns of response on reasoning tasks at adolesence supported Piaget's contention that there are qualitative changes in reasoning at adolescence although, as in other studies, errors in reasoning by adolescents suggested that Piaget overestimated the logical abilities of the formal subject. It is envisaged that further research would be addressed to further investigation of the nature of the changes in linguistic reasoning ability at adolescence and possibly to exploring relationships between the development of inductive and deductive reasoning abilities.

Notes

Footnote 3.1

The error in the denominator is read from the ANOVA table.

Footnote 3.2

The pooled error term is calculated by dividing the pooled sum of squares by the pooled error term:

pooled SS(ERROR)/pooled df(ERROR)

Degrees of freedom associated with the pooled error term are calculated by the Satterthwaite approximation given in Winer, 1971, p.529:

$$f = (u+v)^2/(u^2/f) + (v^2/f)$$

where $u = p(n-1)$MS(subj. within groups)

$v = p(n-1)(q-1)$MS(B X subj. within groups)

$f = p(n-1)$

$f = p(n-1)(q-1)$

REFERENCES

Anderson, J.R. (1976) Language, Memory and Thought Erlbaum: New York.

Anderson, J.R. (1978) 'Arguments concerning representations for mental imagery,' Psychological Review, 85, 249-277.

Anderson, J. R. and Bower G.H. (1973) Human Associative Memory Winston: Washington D.C.

Barratt, B.B. (1975) 'The development of formal reasoning during adolescence', Developmental Psychology, 11, 700-704.

Bart W.M. (1971) 'The factor structure of formal operations.' British Journal of Educational Psychology, 41, 70-77.

Beilin, H. (1975) Studies in the Cognitive Basis of Language Development. New York Academic Press.

Beilin, H. (1980) 'Piaget's Theory: Refinement, Revision or Rejection.' in Kluwe, R.H. and Spada, H. (eds) Developmental Models in Thinking. New York: Academic Press Ch. 11. pp. 245-261.

Bereiter, C., Hidi, S. and Dimitroff, G. (1979) 'Qualitative Changes in Verbal Reasoning during Adolescence during Middle and Late Childhood.' Child Development, 50, 142-151.

Beth, E.W. and Piaget, J. (1966) Mathematical Epistemology and Psychology. Reidel Publishing Company.

Blasi, A. and Hoeffel, E. (1974) 'Adolescence and formal operations.' Human Development, 17, 344-363.

Boden, M. (1979) Piaget. Fontana Paperbacks.

Boole, G. (1854) An Investigation of the Laws of Thought. MacMillan.

Bourne, L.E. and O'Banion, K. (1971) 'Conceptual rule learning and chronological age' Developmental Psychology, ,5, 525-534.

Bracewell, R. J. and Hidi, S.E. (1974) 'The solution of an inferential problem as a function of the stimulus materials.' Quarterly Journal of Experimental Psychology, 26, 480-488.

Bradley, R. and Swartz, N. (1979) Possible Worlds. Blackwell Oxford.

Braine, M.D.S. (1979) 'On the relation between the natural logic of reasoning and standard logic.' Psychological Review, 85, 1-25.

Brainerd, C.J. (1971) 'The development of the proportionality scheme in children and adolescents.' Developmental Psychology, 5, 469-476.

Brainerd, C.J. (1976-77) 'On the Validity of Propositional Logic as a Model of Adolescent Intelligence.' Interchange, 7, 40-45.

Brewer, W.F. and Lichtenstein, E.H. (1975) 'Recall of logical and pragmatic sentences with dichotomous and continuous antonyms.' Memory and Cognition, 3, 315-318.

Brooks, L.R. (1958) 'Spatial and verbal components in the act of recall.' Canadian Journal of Experimental Psychology, 22, 349-368.

Brown, G and Desforges, C. (1979) Piaget's Theory A Psychological Critique. Routledge and Kegan Paul.

Brown, R.W. (1973) A First Language: the Early Stages. London: George Allen and Unwin

Bruner, J. S., Goodnow, J. J. and Austin, G. A. (1956) A Study of Thinking. Wiley

Bruning, J.L. and Kintz, B.L. Computational Handbook of Statistics Scott Foresman and Company.

Bryant, P.E. and Trabasso, T. (1971) 'Transitive inferences and memory in young children', Nature, 232, 456-458.

Bucci, W. (1978) 'The Interpretation of Universal Affirmative Propositions.' Cognition, 6, 55-77.

Bynum, T. W., Thomas, J. A. and Weitz, L. J. (1972) 'Truth-functional Logic in Formal Operational Thinking.' Developmental Psychology, 7, 129-132.

Carpenter, P. A., and Just, M. A. (1975) 'Sentence Comprehension: A psycholinguistic processing model of verification.' Psychological Review, 82, 45-73.

Chapman, L.J. and Chapman, J.P. (1959) 'Atmosphere Effect re-examined.' Journal of Experimental Psychology, 58, 220-226.

Chomsky, N. (1965) Aspects of a Theory of Syntax. Cambridge, Mass. : M.I.T. Press.

Chomsky, N. (1975) Reflections on Language. New York, Pantheon.

Clark, H.H. (1969)'Linguistic processing in deductive reasoning,' Psychological Review, 76, 387-404.

Clark, H.H. and Chase, W.G. (1972) 'On the process of comparing sentences against pictures.' Cognitive Psychology, 3, 472-517

Cohen, J, (1981) 'Can human irrationality be experimentally demonstrated?' The Behavioural and Brain Sciences, 4, 317-331.

Cohen, M. R. (1944) A Preface to Logic Holt.

Cooper, L. A. (1975) 'Mental rotation of random two-dimension shapes.'

Cognitive Psychology, 7, 20-43.

Cresswell, M.J. (1973) _Logics and Languages._ Methuen and Co. Ltd.

Danner, F. and Day, M. C. (1977) 'Eliciting formal operations.' _Child Development,_ 218, 1600-1606.

Donaldson, M. (1963) _A Study of Children's Thinking._ London:Tavistock.

Donaldson, M. (1978) _Children's Minds._ London: Fontana.

Ennis, R.H. (1975) 'Children's ability to handle Piaget's propositional logic: A conceptual critique.' _Review of Educational Psychology_ 45, 1-41.

Ennis, R.H. (1976) 'An alternative to Piaget's conceptualisation of logical competence.' _Child Development,_ _47_ 903-919.

Erickson, J.R. (1974) 'A set analyis theory of behaviour in formal syllogistic reasoning tasks' in Solso, R.L. (ed.) _Theories of Cognitive Psychology._ Erlbaum: New Jersey.

Evans, J. St. B. T. (1978) 'The psychology of deductive reasoning.' in Burton, A. and Radford, J. (eds) _Thinking in Perspective_ London: Methuen.

Evans, J. St. B. T. (1980) 'Current issues in the psychology of reasoning' _British Journal of Psychology,_ 29, 621-635.

Evans, J. St. B. T., (1982) _The Psychology of Deductive Reasoning._ Routledge, Kegan and Paul.

Falmagne, R.J (Ed). (1975) _Reasoning: Representation and Process in Children and Adults._ Hillsdale N.J. Lawrence Erlbaum Associates.

Falmagne, R. J. (1980) 'The development of logical competence: a psycholinguistic perspective.' in Kluwe, R.H. and Spada, H (eds) Developmental Models of Thinking. New York: Academic Press.

Finnochiaro, M.A. (1980) The psychology of logic and the logic of psychology: critique of the psychology of reasoning. in Galileo and the Art of Reasoning. Dordrecht: Reidel.

Flavell, J. H. (1963) The Developmental Psychology of Jean Piaget. London: Van Nostrand.

Flavell, J.H. (1977) Cognitive Development. Englewood Cliffs, N.J.: Prentice Hall.

Flavell, J.H. and Wohlwill, J.F. 'Formal and functional aspects of cognitive development.' in Elkind, D. and Flavell, H. Studies in Cognitive Development: Essays in Honor of Jean Piaget. New York: Oxford University Press, 1969.

Fodor, J.A. (1980) 'Fixation of belief in concept acquisition' in Piatelli-Palmarini M. (ed) Language and Learning: The Debate between Jean Piaget and Noam Chomsky. Cambridge Mass. Harvard University Press.

Fodor, J. (1981) 'The mental representation of quantifiers.' in Peters, S. and Saarinen, E. (eds.) Processes, Beliefs and Questions. Dordrecht:Reidel.

Fodor, J. A. Bever, T.G. and Garratt, M.F. (1974) The Psychology of Language. New York: McGraw-Hill.

Geis, M. and Zwicky, A.M. (1971) 'On invited inferences.' Linguistic Inquiry 2, 561-566.

Gentzen (1964) 'Investigations into logical deduction.' American Philosophical Journal 4, 288-306.

Gilhooly, K.J. and Falconer, W.A. (1974) 'Concrete and abstract terms and relations in testing a rule.' Quarterly Journal of Experimental Psychology, 26, 355-359.

Ginsberg, H. and Opper, S. (1969) Piaget's theory of Intellectual Development: An Introduction Englewood Cliffs, N.J. Prentice Hall.

Gough, P.B. (1966) 'Grammatical transformations and speed of understanding.' Journal of Verbal Learning and Verbal Behaviour 4, 107-111.

Griggs, R. A. and Cox, J.R. (1981) The elusive thematic materials effect in Wason's selection task. British Journal of Psychology, 73, 407-420.

Guyote, M.J. and Sternberg, R.J. (1978) 'A transitive chain theory of syllogistic reasoning.' Cognitive Psychology, 13, 461-525.

Haack, S. (19780 Philosophy of Logics Cambridge University Press.

Handel, S., De Soto, C. and London, M. (1968) 'Reasoning and spatial representations.' Journal of Verbal Learning and Verbal Behaviour, 3, 351-357.

Harman G. (1970) 'Deep structure as logical form.' Synthese, 21.

Harris, R.J. and Monaco, G.E. (1978) 'The psychology of pragmatic implication.' Journal of Psychology, General. 107, 1-31.

Haviland, S.E. and Clark, H.H. (1974) What's new? 'Aquiring new information as a process in comprehension.' Journal of Verbal Learning and Verbal Behaviour, 13, 512-521.

Henle, M. (1962) 'On the relation between logic and thinking.' Psychological Review, 69, 366-378.

Henle, M. (1978) Foreward to Revlin, R and Mayer, R.E. (eds.) Human Reasoning. Washington D.C. Winston.

Hill, S.A. (1961) 'A study of the logical abilities of children.' Dissertation Abstracts 21.

Hunt, E. (1980) 'Intelligence as an information processing concept,' British Journal of Psychology, 71, 449-474.

Hunter, I.M.L. (1957) 'The solving of three term series problems.' British Journal of Psychology, 48, 286-298.

Huttenlocher, J. (1968) 'Constructing spatial images: a strategy in reasoning.' Psychological Review, 75, 550-560.

Inhelder, B. (1962) 'Some aspects of Piaget's genetic approach to cognition.' Child Development,

Inhelder, B. and Piaget, J. (1958) The growth of Logical Thinking from Childhood to Adolescence. Routledge and Kegan Paul.

Inhelder, B. and Piaget, J. (1964) The Early Growth of Logic in the Child. New York

Jackendoff, R. (1972) Semantic Interpretation in Generative Grammar. Cambridge Mass. The MIT Press.

Johnson, M.K., Bransford, J.D. and Solomon, S. (1973) 'Memory for tacit implications of sentences' Journal of Experimental Psychology, 98, 203-205.

Johnson-Laird, P. N. (1975) 'Models of Deduction' in Falmagne, R.J. (ed.) Reasoning: Representation and Process in Children and Adults. Hillsdale, N.J.: Erlbaum.

Johnson-Laird, P. N. (1980) 'Mental Models in Cognitive Science.' Cognitive Science, 4, 71-115.

Johnson-Laird, P.N. (1982) 'Formal Semantics and the Psychology of Meaning.' In Peters, S. and Saarinen (eds.) Processes, Beliefs and Questions 1-68. Dordrecht: Reidel Publishing Company.

Johnson-Laird, P.N. (1982) 'Ninth Bartlett Memorial Lecture: Thinking as a Skill.' Quarterly Journal of Experimental Psychology, 34, 1-29.

Johnson-Laird, P.N. (1983) Mental Models Cambridge University Press.

Johnson-Laird, P. N., Legrenzi, P. and Legrenzi, M. S. (1972) 'Reasoning and a sense of reality.' British Journal of Psychology, 63, 395-400.

Johnson-Laird, P.N. and Tagart, (1969) 'How implication is understood,' American Journal of Psychology, 82, 367-373.

Johnson-Laird, P. N. and Steedman, M. J. (1978) 'The psychology of syllogisms.' Cognitive Psychology, 64-99.

Johnson-Laird, P.N. and Wason, P.C.(1977) Thinking: Readings in Cognitive Science Cambridge University Press.

Just, M.A. (1974) 'Comprehending quantified sentences: The relation between sentence-picture and semantic memory verification.' Cognitive Psychology, 6, 216-236.

Just, M.A. and Carpenter, P.A (1976) 'Comprehension of a negative with quantification.' Journal of Verbal Learning and Verbal Behaviour, 10, 24-253.

Kant, I. (1885) Introduction to logic and essay on the mistaken subtility of the four figures Longman's Green.

Kintsch, W. (1974) The Representation of Meaning in Memory N.J: Erlbaum.

Keenan, J (1978) 'Psychological Issues concerning implication' Journal of Experimental Psychology: General, 107, 23-27.

Kieras, D (1978) 'Beyond pictures and words: alternative information processing models for imagery effects in verbal memory.' Psychological Bulletin, 532-534.

Kirk, R. E. (1968) Experimental Design: Procedures for the Behavioural Sciences. Wadsworth Publishing Company.

Kneale, W. and Kneale, M. (1962) The Development of Logic. Oxford. Clarendon Press.

Knifong, J.D. (1974) 'Logical abilities of young children.' Child Development, 45, 787-83.

Kodroff, J. K. and Roberge, J. J. (1975) 'Conditional reasoning abilities of primary grade children.' Developmental Psychology, 11, 21-28.

Kosslyn, S. M. (1975) 'Information representation in visual images,' Cognitive Psychology, 7, 341-370.

Kuhn, D. (1977) 'Conditional reasoning in children.' Developmental

Psychology, 13, 342-357.

Kuhn, D. and Angelev, J. (1976) 'An experimental study of formal operational thought.' Child Development, 47, 697-706.

Kuhn, D. and Brannock, J. (1977) 'Developmental of the isolation of variables scheme in experimental and "natural experiment" contexts.' Developmental Psychology, 13, 9-14.

Lakoff, G. 'Linguistics and natural logic.' Synthese, 22, 151-127.

Lakoff, G. (1971) 'On generative semantics.' in Steinberg, D and Jakobovits, L. A. (eds.) Semantics Cambridge University Press.

Lefford, A. (1946) 'The influence of emotional subject matter on logical reasoning.' Journal of General Psychology, 34, 127-151.

Lunney, G. H. (1970) Using analysis of variables with a dichotomous dependent variable: an empirical study. Journal of Educational Measurement, 7, 263-269.

Lunzer, E.A. (1965) 'Problems of formal reasoning in test situations.' in P.H. Mussen (ed) 'European Research in Cognitive Development' Monograph of the Society for Research in Child Development, 30, 19-46.

Lunzer, E.A. (1979) 'Formal reasoning: a re-appraisal.' in Floyd, A. (ed) Cognitive Development in the School Years The Open University Press.

Lunzer, E.A., Harrison, C. and Davey, M. (1972) 'The four-card problem and the development of formal reasoning' Quarterly Journal of Experimental Psychology, 24, 326-339.

Lyons, J.(1981) Language, Meaning and Context Fontana Paperbacks.

MacLeod, C. M., Hunt, E. and Matthews, N. N. (1978) 'Individual differences in the verification of sentence-picture relationships.' Journal of Verbal Learning and Verbal Behaviour, 17, 493-507.

Mani, K. and Johnson-Laird, P. N. (1982) 'The mental representation of spatial descriptions.' Memory and Cognition, 10, 181-187.

Manktelow, K. I. and Evans, J. St. B. T. (1979) 'Facilitation of reasoning by realism: Effect or non-effect?' British Journal of Psychology, 70, 477-488.

Marcus, S. L. and Rips, L. J. (1979) 'Conditional reasoning.' Journal of Verbal Learning and Verbal Behaviour, 18, 199-223.

Markman, E. M. (1978) 'Empirical versus logical solutions to part-whole comparison problems concerning classes and collections.' Child Development, 49, 168-177.

Martin, E. (1979) 'The psychological unreality of quantificational semantics.' in Savage, W. (Ed), Minnesota Studies in Philosophy of Science, Volume 9. Minnesota.

Martorano, S.C. (1977) 'A developmental analysis of performance on Piaget's formal operational tasks.' Developmental Psychology, 13, 666-672.

Matthews, N.N., Hunt, E.B. and McLeod, C. M. 'Strategy choice and training in sentence-picture verification.' Journal of Verbal Learning and Verbal Behaviour. 1981.

Mill, J. S. (1874/1961) A System of Logic. Longmans, Green and Co.

Mitchell, D. (1962) An Introduction to Logic Hutchinson University Library.

Monaco, G.E. and Harris, R .J. (1978) Theoretical issues in the psychology of implication.' Journal of Experimental Psychology: General, 107, 28-31.

Morgan, J.J.B. and Morton, J.T. (1944) 'The distortion of syllogistic reasoning produced by personal conviction.' Journal of Society of Psychology, 20, 39-59.

Moshman, D. (1979) 'Development of formal hypothesis-testing ability,' Developmental Psychology, 15, 104-112.

Nagel, E. (1956) Logic without metaphysics. Fre Press.

Neimark, E. D. (1970) 'A preliminary search for formal operational structures.' Journal of Genetic Psychology, 116, 223-232.

Neimark, E. D. (1970) 'Development of formal operational thinking.' in Horowitz, F.D. (eds.) Review of Child Developmental Research, 4. Chicago: University of Chicago Press.

Neimark, E. D. and Slotnick N. S. (1970) 'Development of understanding of logical connectives.' Journal of Educational Psychology, 61, 451-460.

O'Brien, D. P. and Overton, W. F. (1980) 'Conditional reasoning following contradictory evidence: a developmental analysis.' Journal of Experimental Child Psychology, 30, 44-61.

O'Brien, D. P. and Overton, W. F. (1982) 'Conditional reasoning and the competence performance issue: a development analysis of a training task.' Journal of Experimental Child Psychology, 34, 274-290.

O'Brien, T.C. and Shapiro, B. J. (1968) 'The development of logical thinking in children,' American Educational Research Journal, 5, 531-542.

Osherson, D. (1974) Logical abilities in Children. Vol. 2 Underlying Operations. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Osherson, D. (1975) 'Logic and models of logical thinking.' In Falmagne, R. J. (Ed.) Reasoning: Representation and Process in Children and Adults. Hillsdale, N.J.: Erlbaum.

Osherson, D. and Markman, E. (1975) 'Language and the ability to evaluate contradictions and tautologies.' Cognition, 3, 213-226.

Paivio, A. (1975) 'Perceptual comparisons through the mind's eye' Memory and Cognition, 3, 635-647.

Papalia, D. E. and Bielby, D. (1974) 'Cognitive functioning in middle and old age adults.' Human Development, 17, 424-443.

Paris, S. G. (1973) 'Comprehension of language connectives and propositional logical relationships.' Journal of Experimental Child Psychology, 16, 278-291.

Parsons, C. (1960) 'Inhelder and Piaget's 'The Growth of Logical Thinking'.' British Journal of Psychology, 51, 75-84.

Partee, B. H. (1979) 'Semantics - mathematics or psychology' in Bauerle R., Egli, U. and von Stechow, A. (eds.) Semantics from Different Points of View. Berlin: Springer-Verlag.

Peel, E. A. (1967) 'A method for investigating children's understanding of certain logical consequences used in binary propositional thinking.' Britsh Journal of Mathematical and Statistical Psychology, 41, 823-829.

Piaget, J. (1949) Traité de logique. Paris: Libraire Armand Colin.

Piaget, J. (1926) The language and thought of the child. New York: Harcourt Brace.

Piaget, J. (1967) See over

Piaget, J. (1972) The Principles of Genetic Epistemology. Routledge and Kegan Paul.

Piaget, J. (1972) 'Intellectual evolution from adolescence.' Human Development, 15, 1-12.

Piaget, J. and Inhelder, B. (1969) The Psychology of the Child Routledge and Kegan Paul.

Pollard, P. (1982) 'Human reasoning: some possible effects of availability.' Cognition, 12, 65-96.

Pylyshyn, Z. W (1973) 'What the mind's eye tells the mind's brain: a critique of mental imagery.' Psychological Bulletin, 80, 1-24.

Pylyshyn, Z. W (1981) 'The imagery debate: analogue media versus tacit knowledge.' in Black, N. (ed.) Imagery. Cambridge, Mass.: MIT Press.

Quine, W.V.O. (1941/1980) Elementary Logic. Harvard University Press.

Quine, W.V.O. (1951) Mathematical Logic. Cambridge Mass.: Harvard University Press.

Revlin, R. and Leirer, V. O. (1978) 'Understanding quantified categorical expressions.' Memory and Cognition, 815, 447-458.

Revlin, R. and Mayer, R.E. (Eds.) (1978) Human Reasoning. Wiley: New York.

Piaget, J. (1967) Logique formelle et psychologie génétique. In Les Modeles et la Formalization du Compartment (Proceedings of the International Colloquium of the National Center for Scientific Research, Paris, July 5-10, 1965) Paris, 1967, pp. 269-283

Revlis, R. (1975) 'Syllogistic reasoning: logical decisions from a complex data base' in Falmagne, R.J. (ed) Reasoning: Representation and Process. Wiley: New York.

Riegel, K. F. (1973) 'Dialectic operations: the final period of cognitive development.' Human Development, 16, 346-370.

Rips, L. J. and Marcus, S. L. (1977) 'Supposition and the analysis of conditional sentences.' in Just, J. A. and Carpenter, P. A. (Eds.) Cognitive Processes in Comprehension. Hillsdale N. J.: Lawrence Erlbaum Associates.

Roberge, J.J. (1970) 'A study of children's ability to reason with basic principles of deductive reasoning.' American Educational Research Journal, 7, 583-596.

Roberge, J.J. (1976) 'A developmental analysis of 2 formal operational structures, combinatorial thinking and conditional reasoning.' Developmental Psychology, 12, 563-564.

Roberge, J. J. and Paulus, D. H. (1971) 'Developmental patterns for children's class and conditional reasoning abilities.' Developmental Psychology, 4, 191-200.

Roberge, J. J. and Flexer, B. K. (1979) 'Further examination of formal operational reasoning abilities.' Child Development, 50, 478-484.

Roberge, J. J. and Flexer, B. K. (1980) 'Control of variables and propositional reasoning in early adolescence.' Journal of Genetic Psychology, 103, 3-12.

Rumelhart, D.E. (1975) 'Notes on a Schema for Understanding stories.' in Bobrow, D.G. and Collins, A. (eds.) Representation and

Understanding: Studies in Cognitive Science. New York: Academic.

Sachs J.D.S. (1967) 'Recognition memory for syntactic and semantic aspects of connected discourse.' Perception and Psychophysics, 2, 437-442.

Scardamalia, M. (1977) 'Information processing capacity and the problem of horizontal decalage: a demonstration using combinatorial reasoning tasks.' Child Development, 48, 28-37.

Seuren, P (1974) 'Autonoumous versus semantic syntax', in Seuren, P. (ed) Semantic Syntax Oxford University Press.

Shapiro, B.J. and O'Brien, T.C. (1970) 'Logical thinking in children ages six through thirteen' Child Development, 41, 823-829.

Shephard, R.N. (1975) 'Form, formation and transformation of internal representations' in Solso R.L. (ed.) Information Processing and Cognition :the Loyola Symposium. Hillsdale, N.J.: Erlbaum.

Shephard, R.N. and Metzler, J. (1971) 'Mental rotation of three-dimensional objects.' Science, 171, 701-703.

Shine D. and Walsh J. F. (1971) 'Developmental trends in the use of disjunctive concepts.' Psychonomic Science, 23, 171-172.

Smedsland, J. (1970) 'On the circular relation between understanding and logic.' Scandanavian Journal of Psychology, 11, 217-219.

Smedsland, J. (1977) 'Piaget's psychology in practice.' British Journal of Educational Psychology, 47, 1-6.

Smith, C. (1979) 'Children's understanding of natural language hierarchies.' Journal of Experimental Child Psychology, 27, 437-458.

Staudenmayer, H. (1975) 'Understanding conditional reasoning with meaningful propositions' in Falmagne, R.J. Reasoning: Representation and Process. Wiley: New York.

Staudenmayer, H. and Bourne, L. E. (1977) 'Learning to interpret conditional sentences: a developmental study.' Developmental Psychology, 13, 616-623.

Stedmon, J.A. (1982) 'The relationship between logical reasoning and natural language inferences: the class inclusion problem.' in Language, Reasoning and Inference An Interdisiplinary Conference in Cognitive Science. University of Edinburgh, School of Epistemics. Edinburgh, 30th March - 1st April 1982, p.20.

Sternberg, R. J. (1979) 'Developmental patterns in the encoding and combination of logical connectives.' Journal of Experimental Child Psychology, 28, 469-498.

Steinberg, D.D. (1982) Psycholinguistics: Language Mind and World. Longman.

Stone, C. A. and Day, M. C. (1978) 'Levels of availability of a formal operational strategy.' Child Development, 49, 1054-1065.

Strawson P. F. (1971) Introduction to Logical Theory. Methuen and Company Ltd.

Suppes P. and Feldman S.S (1971) 'Young children's comprehension of logical connectives.' Journal of Experimental Child Psychology, 12, 304-317.

Taplin, J. E. (1971) 'Reasoning with conditional sentences.' Journal of Verbal Learning and Verbal Behaviour, 10, 218-225.

Taplin, J. E. and Staudenmayer, H. (1973) 'Interpretation of abstract conditional sentences in deductive reasoning.' Journal of Verbal Learning and Verbal Behaviour, 12, 530-542.

Taplin, J. E., Staudenmayer, H. and Taddanio, J. (1974) 'Developmental changes in conditional reasoning: linguistic or logical?' Journal of Experimental Child Psychology, 17, 360-373.

Tomlinson-Keasey, C. (1972) 'Formal operations in females from eleven to fifty-four years of age.' Development Psychology, 6, 364.

Trabasso, T., Rollins, H. and Shaughnessy, E. (1971) 'Storage and verification stages in processing concepts.' Cognitive Psychology, 2, 239-289.

Van Duyne, P.C. (1974) 'Realism and linguistic complexity in reasoning,' British Journal of Psychology, 65, 59-67.

Van Duyne, P.C. (1976) 'Necessity and contingency in reasoning.' Acta Psychologica, 40, 85-101.

Ward, J. (1972) 'The saga of Butch and Slim.' British Journal of Educational Psychology, 42, 267-289.

Wason, P. C. (1964) 'The effect of self-contradiction on fallacious reasoning.' Quarterly Journal of Experimental Psychology, 16, 3-34.

Wason, P. C. (1968) 'Reasoning about a rule.' , Quarterly Journal of Experimental Psychology, 20, 273-281.

Wason, P. C. (1977) 'The theory of formal operations - a critique.' In Geber, B. (ed.) Piaget and Knowing. London: Routledge.

Wason, P. C. (1977) 'Self-contradictions.' in Johnson-Laird, P. N. and

Wason, P. C. Thinking: Readings in Cognitive Science. Cambridge University Press.

Wason, P. C. and Johnson-Laird, P. N. (1972) Psychology of Reasoning: Structure and Content. Cambridge, Mass.: Harvard University Press; London: Batsford. Wason, P.C. and Shapiro, (1971)

Wildman, T.M. and Fletcher, H.J. (1977) 'Developmental increases and decreases in solution of conditional syllogism problems.' Developmental Psychology, 13, 630-636.

Winer, B.J. (1971) Statistical Principles in Experimental Design. McGraw-Hill Kogakusha Ltd.

Woodworth, R. S. and Sells, S. B. (1935) 'An atmosphere effect in formal syllogistic reasoning.' Journal of Experimental Psychology, 18, 451-460.

Appendix A

Instructions for the shape content problem in experiment 2

"Tim has a game which he and his friends play with lots of different coloured shapes. The idea of the game is to collect as many as you can of three different kinds of coloured shape. These are the shapes Tim was collecting the last time he played...(PICTURE OF SHAPES)...Which of the following sentences that Tim might say about the shapes he was collecting are true and which are false?"

Appendix B

Instructions for the shape content problem in experiment 3

"Tim has a game which he and his friends play with lots of different coloured shapes. The idea of the game is to collect as many as you can of three different kinds of coloured shape. Tim said this about the shapes he was collecting..(Of the shapes that I am collecting all the red shapes are square)..Is this sentence true for these shapes?... these shapes?.."

Appendix C

Instructions for the insect content problem in experiment 4

"Mr. Jones was on holiday in Bugland and while he was there he noticed lots of insects. These are the insects that he saw...(PICTURE).. There were big black insects, big striped insects and small striped insects. When Mr. Jones came home he was telling his friends about the insects he had seen in Bugland. Some of the things Mr. Jones said about the insects were true and some were not true. Can you help by saying which of the following sentences about the insects are true and which are false?

Appendix D

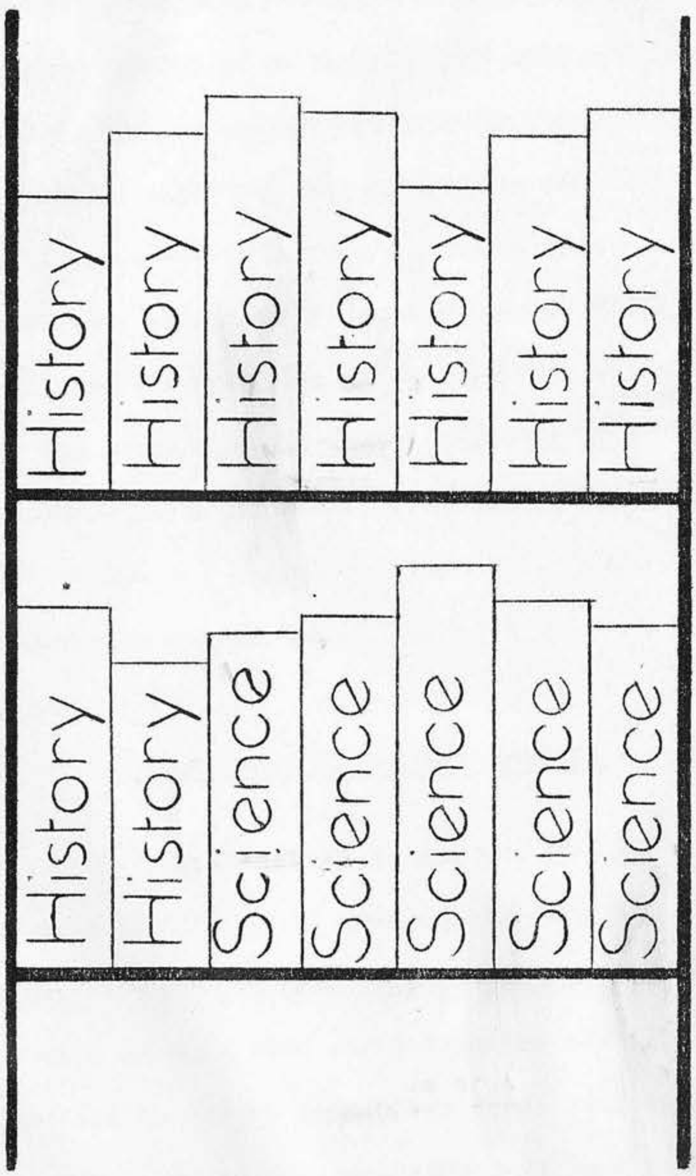
Instructions for the shape content problem in experiment 5

"Tim has a game which he and his friends play with lots of different coloured shapes. The idea of the game is to collect as many as you can of three different kinds of coloured shape. We don't have a picture of the shapes that Tim was collecting the last time he played but we do know that Tim said this about the shapes he wqs collecting .. (CLASS OR CONDITIONAL RULE)....and we know that this is true. Remember Tim was collecting three different kinds of shape and he said about his shapes...(CLASS OR CONDITIONAL RULE). Could this be one of the shapes that Tim was collecting, YES or NO?..How about this one..?..etc."

Appendix E

Story for the shape content problem in experiment 6

"Tim has a game which he and his friends play with lots of different coloured shapes. The idea of the game is to collect as many as you can of three different kinds of coloured shape. We don't have a picture of the shapes that Tim was collecting the last time he played but we do know that Tim said this about the shapes he wqs collecting .. (MAJOR PREMISE)....and we know that this sentence is true. Tim collected three different kinds of coloured shape and as he said (MAJOR PREMISE) he picked up one of his shapes and said (MINOR PREMISE) (CONCLUSION).. YES, NO OR MAYBE."



A) in Professor smart's bookcase all the books on The top shelf are history books

Appendix F

Examples of problems and pictures used in Experiment 7 from different content areas

(A) Book Content

In Professor Smart's bookcase all the books on the top shelf are history books.

Given that you know the rule and you know that:

(A) Book A is on the top shelf.

Is this true? Book A is not a history book.

(A)(1)YES (2)NO (3)MAYBE

(B) Book B is a history book.

Is this true? Book B is on the top shelf.

(A)(1)YES (2)NO (3)MAYBE

etc.

(B) Fish content

In John's fishtank, if a fish is red it has stripes.

Given that you know the rule and you know that:

(A) Fish A is not striped.

Is this true? Fish A is not red.

(A)(1)YES (2)NO (3)MAYBE

(B) Fish B is striped.

Is this true? Fish B is not red.

(A)(1)YES (2)NO (3)MAYBE

etc.

(C) Clock content

In Fred's shop if a clock has a round face it is at six o'clock.

(A) Clock A is not at six o'clock.

Is this true? (A) Clock A has a round face.

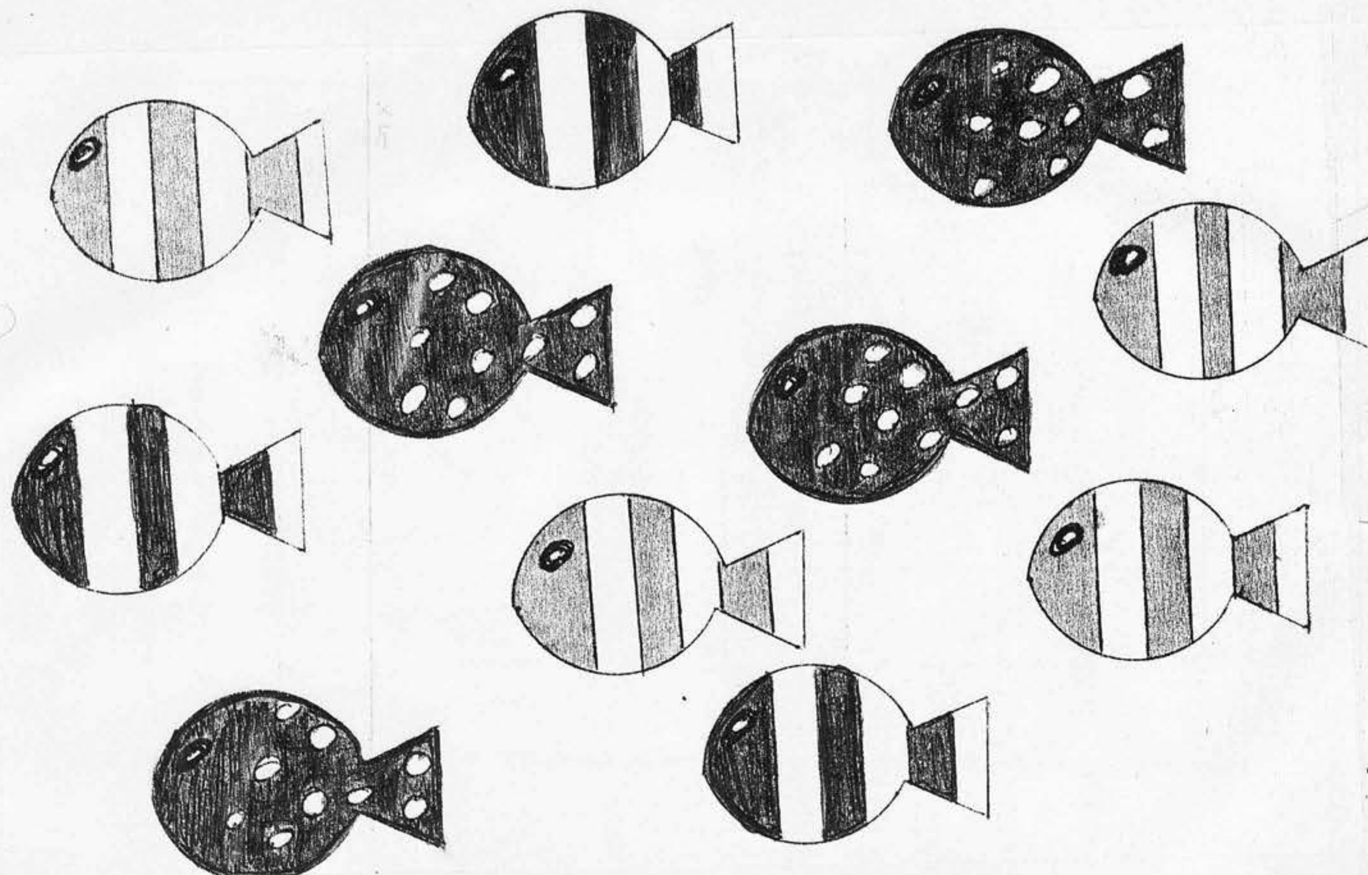
(A)(1)YES (2)NO (3)MAYBE

(B) Clock B has a round face.

Is this true? (B) Clock B is at six o'clock.

(A)(1)YES (2)NO (3)MAYBE

etc.



B) In John's fishtank if a fish is red it has stripes.

grey represents red.
black represents blue.

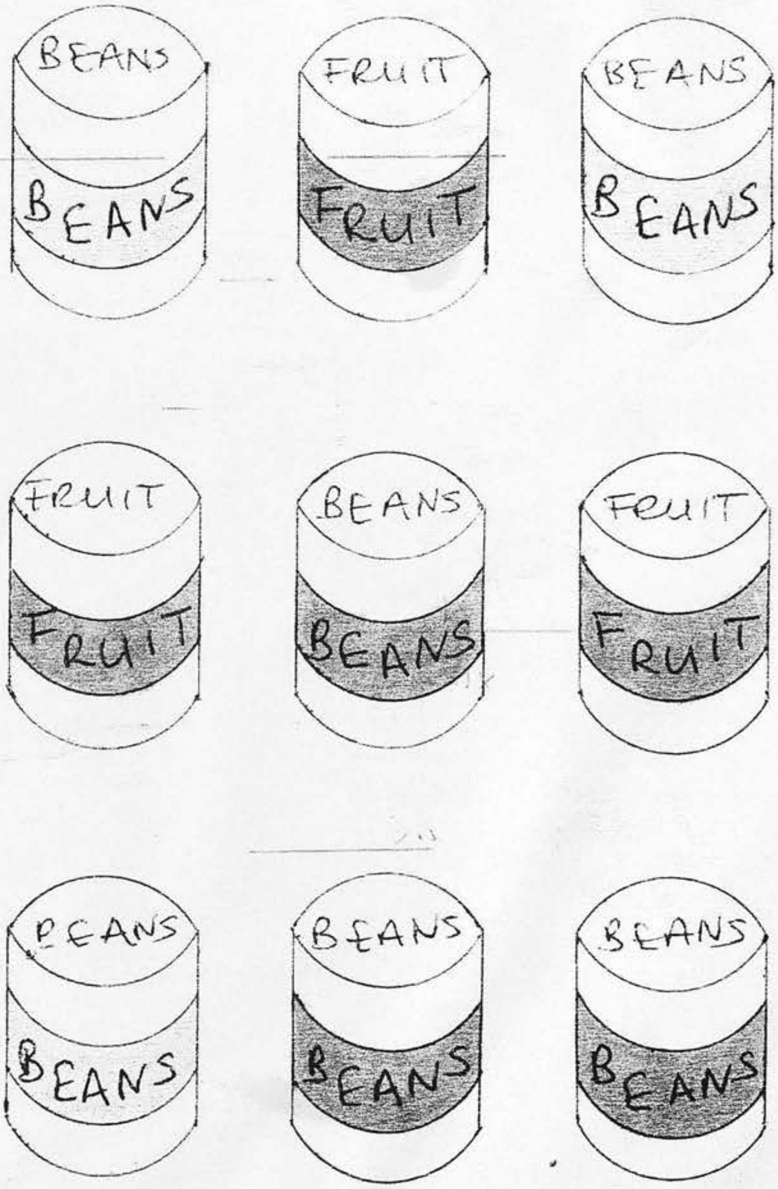
Appendix F cont.

(D) Tins content

In Mrs. Smith's larder all the tins which contain fruit have red labels.

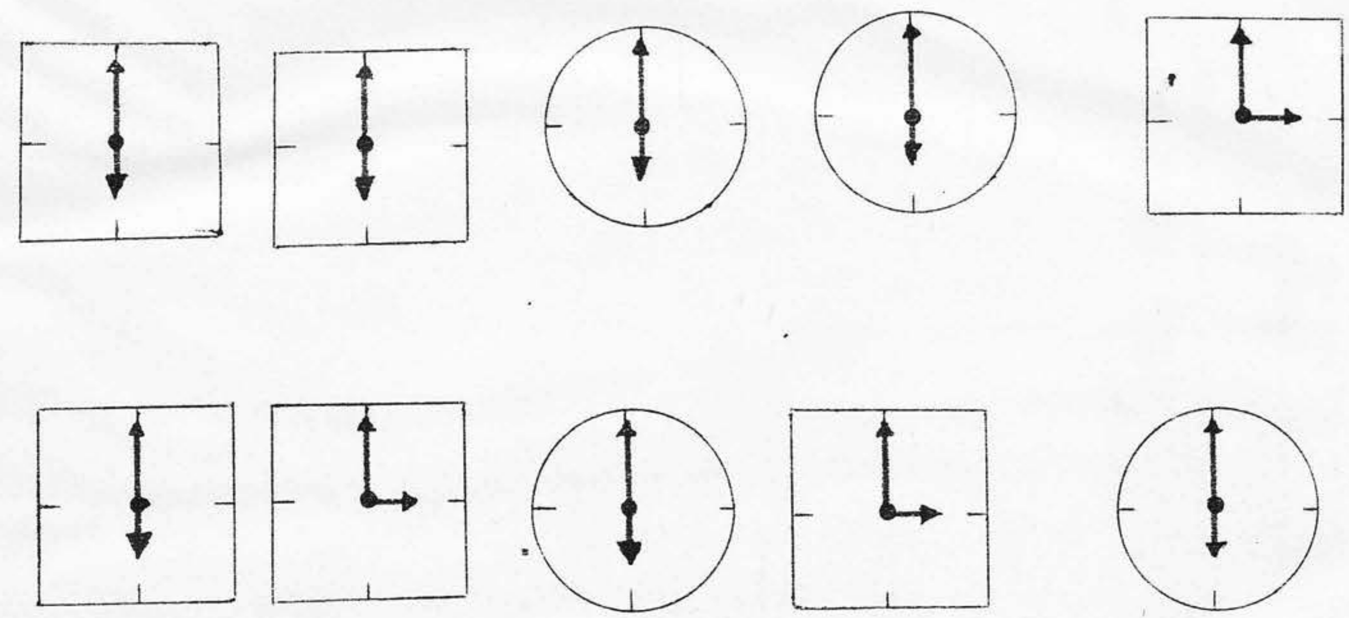
Is this true? (A)Tin A does not contain fruit.
(A)Tin A has a red label.
(A)(1)YES (2)NO (3)MAYBE

Is this true? (B)Tin B does not have a red label.
(B)Tin B contains fruit.
(A)(1)YES (2)NO (3)MAYBE



white labels represent yellow
grey labels represent red.

D) In Mrs. Smiths larder all the tins which contain fruit have red labels



(a) In Fred's shop if a clock has a round face it is at six o'clock.